

# White paper

Type of Report: Final Performance Report and White Paper

NEH Grant Number: HT-251001-16

Project Title: Make *your* edition: models and methods for digital textual scholarship

Project Director: David J. Birnbaum, Professor and Co-Chair, Department of Slavic Languages  
and Literatures, University of Pittsburgh

Grantee Institution: University of Pittsburgh

Date of Submission: 2018-09-30

Submitted by: David J. Birnbaum, Professor and Co-Chair, Department of Slavic Languages and  
Literatures, University of Pittsburgh

# Abstract

This final performance report focuses on the major activities and accomplishments of “Make *your* edition: models and methods for digital textual scholarship,” a three-week Institute in Advanced Topics in the Digital Humanities (IATDH) hosted at the University of Pittsburgh in July 2017. It describes the audiences, the results of evaluations and self-evaluations, the development of sustainable instructional materials, and the continuation of the project and ongoing impact.

## Introduction

The 2017 IATDH “Make *your* edition: models and methods for digital textual scholarship” at the University of Pittsburgh began Monday, July 10 and ended Saturday, July 29. The three-week program included an optional command line and computational groundwork bootcamp with a main program of rigorous instruction on digital methods for philology and for the publication of digital scholarly editions. The Institute presented a cohesive, advanced theory of edition that allowed scholars to build an edition around the individual features of their texts and their research goals, without sacrificing functionality to a homogenizing existing framework.

The digital scholarly edition is more than a reading text with links and annotations. The digital scholarly edition is an integrated platform for performing research, and digital textual scholarship advances as this platform comes to support new types of humanistic inquiry. The principal goal of the Institute was to empower participants who already knew how to mark up their texts (in TEI XML or similarly) to participate directly in the technological conceptualization and implementation of their editions. That level of engagement allows them to undertake philological work informed by an understanding of what is possible technically, and of how to achieve it. The training was designed to anticipate and avert miscommunication or missed opportunity in collaborative situations where no participant in a project fully understands both the textual and the technological issues involved in designing and implementing a digital scholarly edition.

## Description of Institute activities

### Project planning

The instructional team conducted all development inside a GitHub repo (<https://github.com/Pittsburgh-NEH-Institute/Institute-Materials-2017>), which served, through GitHub Pages, as the locus of formatted and published Institute materials (<https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/>). Most instructors participated in teaching all aspects of the Institute, but for development purposes we assigned two instructors to coordinate the preparation of materials for each week: Tara Andrews and Na-Rae Han for Week 1 (bootcamp), David J. Birnbaum and

Ronald Haentjens Dekker for Week 2 (digital editing), and Hugh Cayless and Leif-Jöran Olsson for Week 3 (digital publishing). Senior Institute Assistant Gabi Keane oversaw the preparation of all administrative and logistical materials. Most planning was conducted remotely through email and Skype, except that David and Gabi, who were both based in Pittsburgh, met frequently to coordinate their work, and David traveled to the Netherlands in November 2016, December 2016, and May 2017 to work on Institute preparation in person with Ronald (joined at times by Leif-Jöran).

## Before the Institute

In advance of the Institute, participants were asked to complete a Codecademy Python tutorial (<https://www.codecademy.com/learn/python>). The goal of the tutorial was not to become an expert Python programmer, but to begin to learn how to engage with a programming language. Participants were added to a private discussion forum, which we maintained at the University of Pittsburgh, so that they could ask questions and receive assistance from instructors as they completed the tutorial, but none opted to use this resource, finding the tutorial materials sufficiently clear on their own. The instructional team communicated with participants over email until the GitHub pages site of the Institute repo, introduced above, was sufficiently populated to serve as the primary vehicle for disseminating Institute information. Participants were sent a link to a “Before you arrive” portal page, which provided access to logistical and other preparatory information ([https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/schedule/week\\_1/before\\_you\\_arrive.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/schedule/week_1/before_you_arrive.html)). Participants also received information specifically about housing (<https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/admin/housing.html>) and arrival (<https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/admin/arrival.html>), as well as an orientation to Pittsburgh ([https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/admin/things\\_to\\_do.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/admin/things_to_do.html)), at the GitHub website and through email contact with Gabi.

## Overview of the Institute

Each of the three weeks focused on a different aspect of edition building: computational ground-work skills (“bootcamp”), digital textual editing (“philcamp”), and digital publication (“pubcamp”), with activities in the second and third weeks building on those acquired earlier.

### Principles

The most important principles of the Institute, which permeated all three weeks, were the following: 1) editing is modeling, 2) editing is (participating in) developing, and 3) editing is interfacing.

### Editing is modeling

Digital editions can be modeled on at least two levels: modeling the relationship between the edition products and the research goals and modeling the production process. These can be understood as:

1. *Start with the research objectives in sight.* Digital editions should be designed with research questions, methods, and results in mind from the beginning. Editions are made in different ways and for different purposes, and the editor/developer should think about the research goals before the coding begins.
2. *The digital edition as a pipeline.* The production of a digital edition can be conceptualized as a pipeline, a chain of small operations that contribute incrementally to the end result. The practical advantage of this perspective is that individual steps can be implemented independently of one another, which means that modifications of one step need not require modifying others.

### Editing is (participating in) developing

Digital humanists can acquire the skills to participate meaningfully in the technical development of innovative digital editions. While textual scholars may not do all of their own coding, being able to participate in the technological development improves understanding within the team, and helps distinguish collaboration from compartmentalization.

### Editing is interfacing

Insofar as interfaces can facilitate (or impede) different types of inquiry, the interface and the end-user experience are an expression of the editor's scholarly priorities and judgment. This theme is connected to the idea that an edition is made with a certain research objective in mind: scholars who produce digital editions inevitably make decisions about the types of inquiry their editions will support, and they should make those decisions consciously.

### Bootcamp

Perhaps the greatest practical challenge in any hands-on workshop in the digital humanities is the variation in background knowledge of the computing environment that the participants bring to the class. *Bootcamp: computational groundwork skills*, a novel approach to resolving this problem, dedicated an optional first week to laying a groundwork of computational skills and knowledge that cannot reliably be assumed of digital humanists. Because computational digital humanities evolves rapidly, keeping up with Best Practice is less about learning specific tools than about learning how to apply computational thinking to humanistic inquiry—a competence that is accessible to humanists as long as they are given the opportunity to acquire the skills. For that reason, the bootcamp week focused on familiarizing participants with their computers as tools for managing documents, engaging with such topics as command-line operations; file, program, and operating-system conventions; project management and version control with Git;

introductory programming (illustrated with Python); understanding and using regular expressions; Internet connectivity; Web technologies; understanding and dealing with error messages; and others.

These groundwork skills are essential for engaging creatively with innovative, research-driven digital projects, and while a week of bootcamp training is obviously not enough to become fully comfortable on the command line, it is enough to make substantial progress toward real productivity, and more than enough to demystify the command line and help digital humanists recognize that they are capable of participating actively in developing the technological resources they need to build editions according to their research specifications. Skills were introduced in the context of theory and application, and not just as isolated technical tricks and techniques, to help participants understand how and why they might use a particular tool. Many of the bootcamp activities were split over multiple days to facilitate review (e.g., the second day could review what was covered on the first and then build on top of it) and integration (e.g., regular expressions might be introduced in the context of using *grep* on the command line and then reviewed within Python), and all were put to use in context in the second and third weeks of the Institute.

## Philcamp

*Philcamp: digital philology, creating a digital edition* opened with a discussion of project planning and of modeling as the first step in the development of an edition, inviting participants to consider these foundational issues in the context of their own projects. We introduced the Gothenburg Model of textual collation (GM) not only because collation is a common task in the production of digital editions, but also because GM can serve as an example of modeling edition development as a computational pipeline. Because digital editions may incorporate analytic reports about and graphic visualizations of textual materials, Mike Kestemont (University of Antwerp) joined the instructional team for the end of Week 2 and the beginning of Week 3 to teach the participants about text analytics (reviewing, in the process, many of the bootcamp skills from the first week, including Python). Finally, we introduced XQuery and the eXist-db XML database on Friday to allow participants the weekend to familiarize themselves with the technology for use in the coming week.

The second week was devoted as much to theorizing the creation of digital editions as to learning about specific technologies, and this meant engaging with new ways of examining textual structures and how they can be modeled. We introduced the Layered Markup Annotation Language (LMNL) and Text As Graph (TAG) as a way of helping participants think about their own XML markup decisions in terms of representing their conceptual models of their texts independently of the tacit assumptions underlying prevailing standards and technologies. While XML remains the standard for edition projects because of its maturity, user-base, and infrastructure support, engaging with alternative approaches to text modeling helped participants gain new insights into their own texts and their own XML markup decisions.

## Pubcamp

*Pubcamp: Expressing the edition, digital publication* began with a continuation of Mike Kestemont's instruction to text analytics, focusing on bag-of-words modeling and stylometry. While stylometric analysis is not a traditional part of the (digital) editing process, we consider it in fact to be an essential component of a curriculum for data operations over and computational analysis of text. Next, participants were introduced to publishing frameworks, including simple deployments like GitHub Pages, the CETEIcean toolkit for TEI digital critical editions, and, as a representative of a Python-oriented, API-based web framework, Flask. Participants also worked with XQuery within eXist-db with an eye toward its application to their own editions. At this stage of the Institute, with both theoretical and practical training underway, we invited participants to present lightning talks about their own edition projects, which prompted them to think about (and discuss with others) future directions, project planning, and theoretical approaches to the making of digital editions. In the last two days of this week, participants learned about deploying an *application programming interface* (API) and an eXist-db web app, both as illustrations of concepts involved in planning and publishing a digital edition. Licensing, archiving, and responsible publishing were emphasized in the final days, as most participants had little experience with digital self-publishing.

On the final Saturday the instructional team offered an optional pedagogical review, which explained how and why each week was structured and taught the way it was. Participants asked questions about how they might adapt and reuse materials from the Institute in their own teaching, as many saw a need for that kind of instruction at their home institutions. Instructors facilitated this discussion by offering advice about approaches to edition making, digital editing, and digital philology to those who wanted to focus on adapting Institute materials for training at their own institutions.

## Publication and maintenance

Since the conclusion of the Institute we have revised, expanded, and reindexed our materials to make the entire program more flexible and reusable. Participants and their students can now view tutorials and lessons for nearly all sessions, including those for lessons that were taught completely hands-on during the Institute, without recorded presentation material (live coding). We thought it important to expand the online materials for a number of reasons. First, we wanted to put forward a complete resource, one that would foreground the questions that should be asked as part of the development of a digital edition and offer guidance in the discussion of those questions. Second, we wanted to create a lasting record of the content of the Institute, including teaching materials prepared in advance and post-facto teaching and reference materials derived from presentations. Finally, we found that revising and expanding the material gave us ideas about how we might adapt and improve the materials for reuse. The only exception is that, unfortunately, we are unable to publish some of the teaching materials developed by Mike Kestemont

because of copyright restrictions, but we have included on the Institute GitHub repo two slide sets that he prepared specifically for the Institute.

## Objectives and accomplishments

The four main objectives of the 2017 IATDH were as follows:

1. Empower participants to create editions that are informed by their research goals
2. Make innovative digital theory and methodology developed elsewhere available to digital textual scholars in the US
3. Improve DH pedagogy
  - a. Digital humanists can learn computation and programming
  - b. Task-driven learning can produce focused research outcomes
  - c. Learners can also be teachers
4. Create sustainable, accessible material for teaching, discovery, and reference

### Empower participants to create editions that are informed by their research goals

In our September 2017 survey of participants after the conclusion of the Institute, we learned that many had used their new skills in ways they hadn't considered before. Brian Long "realized the feasibility" of using "digital stylometry for medical texts of uncertain attribution and textual re-use in Greek medical sources" (Appendix V). Gus Riva had already begun using CollateX during editing, and planned to incorporate stylometry in R and NLTK in Python in an upcoming graduate course on digital humanities. Halila Bayramova's doctoral research project now includes "a genetic critical approach" to her in-progress edition of Joyce's *Finnegan's Wake*. Most respondents had recognized new applications of their knowledge by September, and their responses informed our decisions about how to conduct our second, June 2018 survey. Although we originally expected that our evaluation would focus on participant progress in creating and deploying an edition, we had also predicted that the Institute would also have a positive impact on their work outside of edition projects. For this reason, when we solicited feedback in summer 2018 we asked about three main facets of work: editions, general research, and teaching.

The responses, which appear in full in Appendices IV and V, detailed the progress of each participants in those three areas.

Participants at all stages of project planning expressed a particular appreciation for the edition-making instruction, as it informed the creation of their editions at all levels. Bayramova writes that her "edition's model has become much more modular in order to be more ambitious: a consideration has been given to the whole compositional history of *Finnegans Wake* as opposed to only one chapter" (Appendix IV). Her project was in the earliest stages of development when she attended the institute, and she compares its influence to the "chicken-or-egg dilemma...its [the

NEHI's] intellectual impact has blended in too well and now it is hard to tell which bits and pieces have emerged as a direct result of it. (Probably, all of it!)" While Bayramova's Joyce edition was largely in the initial planning stage when she participated in the Institute, Giovanetti's credits the Institute with enabling her to complete her edition of Paolo Bufalini's notebook: "For years the edition hasn't been published because I wanted to do too much with it. We discussed this sort of problems thoroughly during the course and this encouraged me finalizing the work by giving the right limits to the scope of the edition." Even an advanced DH instructor, Elisa Beshero-Bondar, emphasized that "[m]apping out what we want to do with our APIs was by far the most constructive (and most-needed) activity for me. Learning the nuts and bolts of how to do it is something I was already working on prior to the Institute in my work with eXist-db, but my view of what was possible was expanded by consulting the expertise of the instructors."

General research skills can be difficult to gauge, particularly as research evolves throughout a project. Here we will not consider whether or not participants were successful at mastering command line or Python simply because we never expected them to do so. Instead, we focus on what Elli Bleeker called "a 'computational mindset'... which is highly valuable for any kind of research yet not something one usually learns from your average humanities program. For me, the true value of this Institute is twofold: first, acquiring a computational way of thinking and secondly, learning how to apply that way of thinking to scholarly editing." We saw most participants engage with these ideas in their responses, as they outlined which technologies they were using, how and why they were using them, and what they planned to do next in their projects. The responses as a whole reflect the modular approach to research outlined in the Institute. Ohge writes "...the NEH Institute has most directly affected my most recent digital project, an edition of Mary Anne Rawson's 1834 anti-slavery anthology *The Bow in the Cloud*. This edition is not only creating a versioning text based on the original manuscripts of the anthology, but it also will include a network analysis and visualisations of variants, most frequent terms, & sentiment words. I'm also using eXist to pipe my files into a graph database (for the network analysis). All of those components were clearly facilitated by the additional training at the NEHI. The specific technical steps Ohge outlines are part of the "computational mindset" Bleeker describes; not only is Ohge considering the implications of the technical activities he describes, but he understands how he can implement them himself.

The Institute's approach to making teachers out of participants was reflected in many ways in the responses, from workshops to collaborative projects to code clubs. Andrea Nichols wrote "I am finishing up a syllabus for a Macroanalysis course (based largely on what I learned from the Institute), and I want to write a few outlines or guides for running an afternoon workshop or Hack-a-thon. There seems to be a growing need or desire to do these quick training sessions." She emphasizes the usefulness of these "transcribe-a-thons and hack-a-thons" at conferences, with "female code clubs," and at "local universities without DH faculty." Christopher Ohge made a more direct link: "I am including several lessons from NEHI: in my July 2018 London Rare Books



School course on digital editing (<https://cmohge1.github.io/lrbs-digital-scholarly-editing/>), I am including sessions on digital text modelling, collateX, & lmn1, for example. In workshops for the UL School of Advanced Study's digital humanities initiative (<https://www.sas.ac.uk/projects-and-initiatives/digital-humanities/>), I have also recently included command line & regular expression bootcamps based on the NEHI." Most consistently, participants have told us that the most instrumental contribution to the continuation of their teaching is the GitHub Pages site. Susanna Alles Torrent writes "I must say that I really appreciate the organization of the Institute under the form of a website, and especially that we can still access to all the instructor's materials. I use it quite often for my work."

## Make innovative digital theory and methodology developed elsewhere available to digital textual scholars in the US

While it remains true that funding (and therefore training opportunities) for DH scholarship in the EU is more readily available than in the US, the Institute was able to provide essential training to US-based participants that is simply not available elsewhere. This objective is bound to the goals that "participants will become teachers" and that our materials should be available into the future.

## Improve DH pedagogy

Our emphasis on teaching strategies for learning to use technological tools, rather than merely "tooling up", showed digital humanists that they can learn (and teach) computation and programming in a useful way without a major investment of time or funding. This concept was especially useful in the bootcamp, when participants were most likely to wonder, "When am I ever going to use this?" That's a common and fair question for humanists to ask, as their main focus should always be their research question. The Institute's integration of practical computational skills and methods into the expression of research and scholarship helped participants to draw connections between new concepts and familiar skills, thus increasing their comfort with new material and prompting them to consider novel approaches to research questions.

We integrated computational and editorial methods during all three weeks by using task-driven activities called *code labs*, which encouraged active participation and open dialogue with instructors. When a participant hit a snag, they could call on instructors, all of whom attended nearly every session, for individual assistance, and they learned a sustainable approach to troubleshooting that allowed them to continue coding after the workshop. Debugging became a community-oriented task (much as it often is in the real world), which eased participant anxiety about errors.

Interaction in the classroom, during breaks, in their lodgings, at meals, and elsewhere created opportunities for participants to learn together and to teach one another. Impromptu code workshops often sprung up after hours, as the intellectually diverse group returned from social outings

to debug and work through coding challenges together. In the time since the Institute ended, many participants integrated course materials into their classes at home institutions, shared their discoveries with project teams, and recommended our materials to others.

## Create sustainable, accessible material for teaching, discovery, and reference

We chose to keep our Institute materials in a GitHub repository for easy collaboration and distribution during preparation and the three weeks we were in session. This strategy helped us integrate teaching GitHub to participants, as well. After the end of the Institute, GitHub became even more crucial as instructors and the senior assistant continued to edit materials remotely. Real-time updates to materials could be tracked, and therefore reversed or enhanced, by anyone on the team. One participant sent us a pull request with suggested changes to a tutorial. The GitHub Pages feature made presentation cohesive and simple.

To make materials more accessible to non-participants, we have expanded many of our tutorials, added definitions, and linked to many outside resources. One no longer needs the context of the Institute or verbal instruction to use the materials, which makes them adaptable for teaching and edition-making alike.

To ensure that the materials continue to be available, the entire GitHub repo is mirrored by the University of Pittsburgh Library System (ULS). We link to the materials on both GitHub and ULS in the Grant products section, below.

## Audiences

Our 2017 IATDH received 33 applications, of which we accepted 26: 9 graduate students, 2 post-doctoral fellows, 10 faculty, and 5 non-faculty researchers. Of the 26 accepted participants, 19 were studying or working in the US, 5 were based in Europe, 1 was from South America, and 1 was from Asia. Three persons withdrew just before the beginning of the Institute for personal reasons (1 graduate student at a US university had visa complications, 1 graduate student from Europe had a sudden illness in the family, and 1 faculty member from the US was asked to make unexpected and substantial changes in a book manuscript facing a publication deadline). This left us with a total of 23 enrolled participants. Participants were humanists and humanistic social scientists whose projects were document-based.

The archived materials will be useful to a similar audience, as well as to those with different backgrounds and interests than our participants. These include:

1. Researchers who know they might need to use digital methods, but are as yet unsure of which methods to use, or of *how* they might learn to use them without a major investment of time or funding.

2. Developers who are too advanced for a comprehensive three-week course on digital editions, but who are not yet familiar with specific topics, such as the Gothenburg Model of textual collation.
3. Digital humanists who are not interested specifically in edition making, but who seek technological guidance, such as command line skills or XQuery tutorials.
4. Scholars interested in thinking in innovative ways about the theory underlying their editions, as the Institute puts forward a dynamic, flexible concept of edition-making that prioritizes project-specific research questions and outcome goals.

Those wishing to teach any topics covered during the Institute may find the pedagogical review (<https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/pedagogy.html/>) helpful. Participants expressed a desire to remix and reuse materials in part in their own courses, including Christopher Ohge's July 2018 London Rare Books School course on digital editing (<https://cmohge1.github.io/lrbs-digital-scholarly-editing/>). Participants and non-participants alike may find the materials on the site useful as lessons or assignments, as many tutorials can be adapted or expanded easily. All site pages are written in Markdown, which can be changed to plain text, HTML, or other formats (within GPL 3.0 guidelines).

Finally, we intend that the online materials will serve as a reference source for those building and learning on their own. The stability of the site, availability through an open GitHub repository, and the range and depth of coverage make it a reliable and useful reference for humanists in all stages of project development. By linking to outside sources and tutorials, the site also introduces new researchers to the broader community of thought on these subjects.

## Evaluation

### Overview

The most meaningful assessment of the success of the Institute is the practical impact it has on the participants' ability to design, develop, and deploy digital editions that answer their individual research needs. As described in our original proposal, in order to provide time for the participants to implement what they had learned, we solicited feedback approximately one year after the conclusion of the Institute, in June 2018. Our assessment of the Institute is based primarily on participant responses to this inquiry, with supplementary information from three other sources: 1) responses to a preliminary inquiry circulated in September 2017, 2) daily brief feedback during the Institute (which enabled us to make on-the-fly adjustments to the curriculum), and 3) conversations and correspondence with the participants during and after the conclusion of the Institute. All paper responses (to preliminary and final email inquiries, as well as daily feedback) are included in Appendices IV and V.

In conversation with the instructors and in their daily feedback, the participants identified the following as particular strengths of the organization and administration of the Institute:

1. *One-on-one access to instructors.* We designed the Institute so that all six principal instructors, plus the two assistants, would be present and circulating during the sessions, with the goal of making individualized help available during in-class *code labs*. The instructors ate lunch with the participants every day, and most stayed with them in the same dormitory accommodations.
2. *Social contact.* Participants stayed in the same dormitory accommodations and met frequently in large or small groups for meals, coding sessions, and informal social excursions. This helped build a sense of community that extended into the Institute sessions, where participants looked out for and helped one another as needed. Gus Riva writes that “Seeing projects from other colleagues and talking about them was very helpful to redefine my own goals and assumptions on digital editing.”
3. *Flexibility.* Daily feedback surveys enabled the instructors to modify the curriculum on the fly in situations where participant responses indicated that the original planning had been overly ambitious or optimistic or when participants felt the desire to spend more time on a particular topic than the original schedule allowed. One especially popular event was an additional eXist-db and XQuery evening workshop organized and conducted by Leif-Jöran supported by a few other instructors in response to participant request.

One of the most innovative components of the Institute was the week-long *bootcamp* module, designed to equip participants with the basic command-line literacy that is needed to escape the limitations of web interfaces and canned software packages. Several participants wrote, in response to our September 2017 request for feedback, that they began applying many of the bootcamp skills immediately to their own research (Appendix V). These included batch processing at the command line (Marie-Claire Beaulieu), regular expressions for data cleaning (Beaulieu), Git and GitHub for project management (Halila Bayramova), Python (Riva writes about how Python with `xmltree` provided a new perspective on XSLT; Beaulieu writes that learning to read Python let her figure out how to begin to read other languages), collation and alignment (Riva, Chris Ohge, Brian Long).

The intellectual core of the Institute was the idea that the preparation and publication of a digital edition could be understood and managed as a *pipeline*. This aspect of computational thinking permeated all aspects of the Institute, from literal Unix command-line pipelines during the bootcamp to conceptualizing entire edition projects as developmental pipelines. Participants identified pipelining as one of the most valuable takeaways from the Institute; for example, Ming Yeung Cheung cited the pipeline model as a focal area in a training workshop he was preparing for his colleagues, and Ohge said that “echoing what I learned from David Birnbaum, I would

say that digital editing is the computational pipeline (from encoding and processing digital documents to interface design) that brings a scholarly edition to a digital medium”.<sup>1</sup>

With respect to the theory and practice of developing and deploying digital editions, which was the focus of the second and third weeks of the Institute, participants singled out both technological training (Riva highlights collation; he and Ohge foreground stylometry) and non-technical features that were part of the Institute curriculum. Those non-technical topics included licensing and publication (mentioned by Bayramova); designing with user profiles in mind from the beginning of the project, not only with respect to deployment (also Bayramova); the documentation of both software and data (Paul Hackett); LOD (Francesca Giovannetti); API design (Cheung); and visualization theory and practice for both research and public access (Hackett).

The Institute had both an immediate pedagogical goal (teach the participants to design and implement digital editions by applying computational thinking to their research) and a meta-pedagogical one (teach the participants how to train others). The final half-day pedagogical review was designed to support that second goal by providing an explanation of how and why we constructed the curriculum as we did. By exploring the pedagogical rationale and methodology of the Institute together, we intended that participants would be prepared to design and conduct similar training workshops for colleagues at their home institutions. The impact of this aspect of the Institute turns out to have been substantial. Riva talked about his plans to teach DH with Python/NLTK and R/Style. Les Harrison wrote about teaching an “Editing the critical edition” course. Fernando Nascimento (who is a professor of computer science) wrote about incorporating XML, stylometry, and Python into his “Computation in context” and “Digital textual analysis” courses. Ohge told us about computational pipelines, the philosophy of markup, and visualization as part of edition in his teaching. Rikk Mulligan wrote about the different pedagogical roles of short workshop sessions, day-long workshops, and longer mini-courses. Hackett, though not teaching DH immediately after the institute, wrote of integrating a broad overview of the history, theory, and practice of DH into both digital edition and general curricular planning.

The Institute fostered new opportunities for networking for both participants and instructors. Elisa Beshero-Bondar and David J. Birnbaum (Institute Director) taught XPath (including XSLT, XQuery, and Schematron) at the DHSI in summer 2018, with Hackett and Ohge among the registered participants. Albertina Walker-Hughey arranged for Hackett to advise her about API design and implementation for her edition projects.

## Challenges and solutions

An early and unexpected challenge was that one of the original core developers of the Institute proposal, Joris van Zundert (Huygens Institute for the History of the Netherlands, Royal

---

<sup>1</sup> <https://englishstudies.blogs.sas.ac.uk/2017/11/03/an-interview-with-dr-christopher-ohge/>

Netherlands Academy of Arts and Sciences), had to withdraw from the Institute planning and teaching for personal reasons. Fortunately, we were able to replace Joris with two highly qualified colleagues: Na-Rae Han (University of Pittsburgh), who took over Joris's duties in the bootcamp week, and Hugh Cayless (Duke University), who replaced Joris in the weeks devoted to digital editing and digital publishing.

During the weeks that the Institute was in session we identified three challenging sets of issues: 1) software installation difficulties; 2) uncertainty about goals; 3) accommodating diverse participant expectations.

### Software installation

Software installation is a constant challenge in computational DH training, and it was compounded in our case by the variety of tools we introduced and our commitment to supporting multiple versions of Windows, MacOS, and GNU/Linux. The instructors and assistants addressed this challenge by splitting the participants into two separate GNU/Linux+MacOS and Windows groups for the first sessions, and by working one-on-one with the participants, and we adjusted the curriculum on the fly in situations where installation difficulties caused us to fall behind our original schedule. Although these methods proved successful, in future Institutes we would budget additional time for installation.

We also continue to explore the possibility of deploying more Institute software in containers (we use Docker for this purpose). Docker itself can be difficult to install in a heterogeneous environment like ours, where challenges included old hardware and software (we found several participants running operating systems not updated since 2010), new hardware that did not support Docker's virtualization requirements, and Docker assumptions about underlying file systems. Once we help the participants through that installation process, with some practice in looking up error messages, deploying containerized applications is straightforward. We now make eXist-db, CollateX, and Alexandria available inside Docker containers.

### How am I going to use this?

The bootcamp provided training in core computational skills that are needed by anyone who aims to participate in the development of project-specific tools. These topics include command-line operations; file, program, and operating-system conventions; project management and version control with Git; introductory programming (illustrated with Python); Internet connectivity; Web technologies; etc. Bootcamp topics were introduced as skills that were going to be contextualized in the second and third weeks of the Institute, and, indeed, by the end of the Institute participants who had not previously known that there was such a thing as a command line were operating on it comfortably, at least in some contexts. For example, we could begin a session with an instruction like “pull and merge the changes from the master branch of the upstream repo” and people could perform that action without assistance. But although the bootcamp

generally met its goals, its stand-alone nature, with a promise that the uses of the various skills would become clearer later in the Institute, led to moments of confusion or frustration along the lines of “I don’t understand why I’m learning this.”

The provision of an optional, stand-alone bootcamp that focused on techniques that would be applied only in the second and third weeks of the Institute was based on our observation, in prior workshops, that DH learners seemed to sort themselves into two groups: those who already have basic command-line literacy and those who do not. As it turned out, though, almost all participants opted into the bootcamp, and even those who already had meaningful command-line experience told us that they found value in the content. Those participants found this week useful as review, because there were details they didn’t know previously, because they appreciated learning how to teach command-line skills, or because of the opportunity to share their knowledge with others during hands-on code-lab sessions. Based on that observation, in future Institutes we would recommend integrating the bootcamp skills into the rest of the activities throughout, so that new techniques could be practiced *in context* immediately, instead of introducing them in a separate, non-obligatory preparatory module. This does not mean that stand-alone bootcamp might not also be valuable, but as part of a longer Institute with a specific set of coordinated outcome goals (editing and publishing digital editions, in our case), we now think that we could teach those skills more effectively by integrating them closely with the other outcome goals rather than only building on them in later modules and reviews.

### Responding to diverse expectations

The technologies employed in digital textual editing and publishing, along with the community’s understanding of Best Practice, are in constant and rapid motion, which means that the most valuable training an instructor can provide comes not as a set of answers, but as a set of methods for researching and discovering answers on one’s own. The proverbial model for this approach distinguishes giving someone a fish and feeding him for a day versus teaching him to fish and feeding him for the rest of his life. In computational humanities, learning to fish means learning how to Look Things Up. Some learners felt frustrated when told to look something up, but one participant highlighted the importance of this type of training when she wrote: “I especially appreciated the instructors’ advice on how to search and find answers to our questions rather than providing us with the answers themselves.” (Bayramova). In response to both types of learners, we guided those unwilling to Look Things Up through our own search strategies. In situations where an instructor or other participant might have had an ready solution to a problem, we encouraged them to guide the participant through the research process instead of just giving them the answer. By the end of the Institute, all participants could (though sometimes a bit reluctantly) begin searching for answers themselves before approaching someone else with their question. Not only does this often lead to answers, but it also prepared the participants to ask questions on discussion platforms like Stack Overflow, where researching an issue before posting a question is part of the expected professional etiquette.

## Continuation of the project

Participants who responded to the one-year call for feedback reported varying degrees of completeness and complexity of their digital projects. This outcome is to be expected from such a diverse group of participants—some came into the Institute with nearly completed projects, while others had only an idea of what they wanted to do. As instructors and facilitators, our task in the past year has been to update and improve the online materials available to participants and their students in an effort to keep those with projects in progress on the right track.

In coming years, we will continue to mediate participant hurdles through the established networks created during the Institute. As digital projects reach maturity or change form completely, we will update the “Editions by Institute Participants” and “Selected Exemplary Editions” pages. By creating sustainable, easy-to-update materials (with easily tracked changes, through our site’s GitHub rep), the progress of Institute goals will not end with the close of the grant. Some participants suggested they might have benefited from webinars or podcasts on our topic, but we felt those media would remove the essential community and interactivity of our sessions. As Brian Long points out, “[in] the few cases where instruction progressed a bit too rapidly, the accessibility and helpfulness of the instructors proved to be a significant asset” (Appendix IV). Often, debugging and troubleshooting require physical access to the user’s computer, which is one reason that in-person lessons are an essential part of this type of training.

In response to some of the suggestions offered by participants, we decided to apply for a 2019 IATDH Grant in the 2018 competition, to be held in Summer 2019. With the experience of the 2017 IATDH behind us, we reevaluated whether to apply again for a three-week program, as it can be difficult for participants to commit to such a long period. In the end, however, we concluded that the combination of broad perspective and hands-on experience could not be managed in less than three weeks, and we further decided to make all three weeks mandatory for all applicants. The decision to make the bootcamp content mandatory stems is rooted in the fact that almost all participants enrolled in it, and even experienced coders like Elisa Beshero-Bondar identified it as valuable. But some participants also observed, especially in their daily feedback, that the application of bootcamp activities was sometimes unclear without the larger editing and publishing tasks that were introduced only beginning the second week. This led us to propose, for our next IATDH, that the bootcamp activities and topics be integrated into the application context, so as to contextualize these essential skills from their first introduction.

## Grant products

- All materials on GitHub at <https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/>, mirrored by University of Pittsburgh University Library System (ULS) at <http://d-scholarship.pitt.edu/35394/>



## Long-term impact

### Bootcamp

In our original proposal, we suggested a command-line and programming bootcamp for humanists to fill what we saw as a need in the digital humanities community. Many scholars come across digital methods and see a great opportunity, but lack the basic computer skills to make full use of those tools. We wanted to enter the more complex phases of the Institute with a group of people who had those skills: they were confident in their own abilities to manage files, run commands, and program at a basic level. But more than that, we wanted to fill that gap, which leaves some digital humanists behind. Many feel intimidated by their computers, as they don't understand how they work; they don't know what they don't know; and they are afraid of making mistakes. Without training of the sort provided by the Institute, many regard computers as appliances for web browsing, word processing, and email. The bootcamp seeks to resolve the knowledge divide and level the playing field for DH researchers.

The novel bootcamp model fills the DH community's need for instruction in basic computational groundwork skills. The Perseus Project used this model in their own 2018 NEH Institute proposal,<sup>2</sup> where, during their bootcamp week, "participants follow a common curriculum and develop a shared set of skills" (Perseus). Their NEH application specifically cites our 2017 Institute, writing that our model "will inform the work of the proposed Institute." Although we integrated the bootcamp with the other materials in our subsequent 2019 IATDH proposal, we nonetheless found the bootcamp successful. As mentioned above, nearly all the participants attended, and those who did not likely would have benefited from it in some capacity. Our integration of the bootcamp into other activities in our new proposal reflects our desire to address the biggest question participants had during the first week, "How am I going to use this?" head on, rather than with reassurances that they would need it later.

### Pedagogical review

In an effort to encourage participants to go back to their home institutions and evangelize digital methods, we organized a Saturday morning review of why the Institute was structured as it was. The overall goals and roadmap were explained, followed by explanations of specific teaching methods for each topic. Participants asked questions about how they might adapt certain materials to their own classrooms, integrate certain technologies into their existing project teams, and design courses around what they had learned.

The Saturday sessions also allowed one last networking opportunity, as participants were able to discuss plans for the future in a more relaxed classroom environment. Participants were also able

---

<sup>2</sup> "Digital editions, digital corpora and new possibilities for the humanities in the academy and beyond."

to offer instructors feedback in an open, direct way, which in turn helped participants think about how they might teach material they had struggled with as learners.

## Conceptualization and modeling

Instead of teaching participants to “tool up” on specific technologies, the Institute emphasized conceptualization and modeling of editions. Designing a project to fit the requirements of a particular tool can be limiting because the tool can constrain the research possibilities, rather than letting the research goals dictate the choice of tools. For this reason, scholars should be able to formulate a research question, conceptualize their edition in relation to that question, and then implement and deploy that edition.

The Institute approached the problem of “tooling up” by using contemporary tools not as end goals, but as examples of edition making as a computational pipeline. In our examples, we used the Gothenburg model to show the many different processing steps necessary for collation. For a scholar not interested in collation, the tools themselves might not have been interesting, but the approach modeled was still ideal. In the future, many of the technologies we talk about in our materials may fall out of use, but the skills learned in acquiring them (read error messages, Look Stuff Up, plan a pipeline to meet specific research goals) will still apply.

## General advice

### Before the program

- Assemble the instructional team, including assistants, and allocate supervisory responsibility for different components of the Institute
- While developing instructional materials in a way that is consistent with the goal of *sustainable pedagogy*, keep the focus on learning how to learn, and not just on learning fixed skills
- Undertake development in a framework that supports publication and use in teaching. In our case, we developed in GitHub and published through GitHub Pages.
- Advertise the Institute widely. Select the participants with an eye toward diversity of interests, experiences, and backgrounds.

### During the program

- House participants and instructors in the same building, if possible. After hours collaboration and debugging happen in this kind of open, cooperative community.
- “Elving” (additional instructors who wander among and help participants during hands-on activities) is important. Instructors who aren’t lecturing should walk the floor during code sessions to resolve individual problems without interrupting the session.

- Allow participants to signal for help, and for their availability to help others. In our sessions, a pink sticky note on one's laptop indicated a need for help, while a yellow sticky note indicated that a participant had completed the task and was available to help others.
- Devise a way to gauge the depth of participant understanding. Asking “Any questions?” or “Do you understand?” doesn't work; it's more effective to ask participants to complete a task that requires them to have understood what was just taught.
- Build review into the instructional time. Proficiency with technology requires hands-on practice, and review allows instructors to identify and respond to patterns of misunderstanding and adapt the lesson material to participant needs.

### After the program

- The director should reach out to participants within a month of the end of the Institute. In our case the early response rate was about 50%, but the feedback they provided was of very high quality, and helpful in our revision of Institute materials for publication. This early outreach also communicates that instructors are available for future consultation.
- To give participants adequate time to begin to implement what they have learned in the Institute, reach out to them again for a final evaluation approximately one year after the conclusion of the Institute sessions. The best measure of the lasting effects of the Institute is how the content has influenced the research and teaching of the participants.
- Keep track of the research interests and projects of participants. In the future, instructors may come across another researcher, topic, or technology they can connect to former participants. In our case, participants and instructors have built on acquaintances made at the Institute to collaborate in development and teaching projects.

## Appendices of supporting materials

### Appendix I: Advertising the call for applications

The Call for Applications was published on Humanist (<http://dhhumanist.org/>), Digital Classicist (<http://www.digitalclassicist.org/>), Digital Medievalist (<https://digitalmedievalist.wordpress.com/>), TEI-L (<http://www.tei-c.org/Support/index.xml#tei-l>), WWP-Encoding (<http://listserv.neu.edu/cgi-bin/wa?A0=wwp-encoding>), and DHUF Digital Humanities Flanders ([dh\\_flanders@googlegroups.com](mailto:dh_flanders@googlegroups.com)). The text of the advertisement is available at <https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/admin/call.html>.

## Appendix II: List of participants with affiliations

Our 2017 IATDH received 33 applications, of which we accepted 26: 9 graduate students, 2 post-doctoral fellows, 10 faculty, and 5 non-faculty researchers. Of the 26 accepted participants, 19 were studying or working in the US, 5 were based in Europe, 1 was from South America, and 1 was from Asia. Three persons withdrew just before the beginning of the Institute for personal reasons (1 graduate student at a US university had visa complications, 1 graduate student from Europe had a sudden illness in the family, and 1 faculty member from the US was asked to make unexpected and substantial changes in a book manuscript facing a publication deadline). This left us with a total of 23 enrolled participants.

1. Allés Torrent, Susanna; University of Miami
2. Bayramova, Halila; Trinity College Dublin
3. Beaulieu, Marie-Claire; Tufts University
4. Beshero-Bondar, Elisa; University of Pittsburgh, Greensburg
5. Bleeker, Elli; Huygens ING, Royal Netherlands Academy of Arts and Sciences
6. Cheung, Ming Yeung; Fu Jen Catholic University, Taiwan, and  
Divinity School of Chung Chi College, Chinese University of Hong Kong
7. Eberle, Roxanne; University of Georgia
8. Fernández Riva, Gustavo; University of Buenos Aires
9. Giovannetti, Francesca; Consortium GARR
10. Hackett, Paul; Columbia University
11. Harrison, Les; Virginia Commonwealth University
12. Hegde, Krishna; University of Pittsburgh
13. Long, Brian; Pontifical Institute of Medieval Studies
14. Magni, Isabella; Center for Renaissance Studies, Newberry Research Library
15. Mulligan, Rikk; Carnegie Mellon University
16. Nascimento, Fernando L.; Bowdoin College
17. Nichols, Andrea; University of Nebraska, Lincoln
18. Ohge, Christopher; University of California, Berkeley and University of London
19. Parker, Rebecca; Loyola University, Chicago
20. Reggiani, Nicola; University of Parma
21. Walker-Hughey, Albertina L.; Texas Southern University

22. Williamson, Elizabeth R.; Folger Library and University of Exeter
23. Zimmer, Mary Erica; Boston University

## Appendix III: Original application narrative

### A. Intellectual significance

The digital scholarly edition is much more than a reading text with links and annotations. The digital edition is a *workstation*, an integrated platform for performing scholarly research, and much as philology advances with new theories and methods, digital textual scholarship advances as the workstation comes to support new types of research questions. By training philologists to participate directly in the technological conceptualization and implementation of their editions, the Institute will empower them to undertake philological work that is informed by an understanding of what is possible technically, and of how to achieve it. This training responds to the serious risk of miscommunication or missed opportunity in collaborative situations where no participant in a project understands fully both the textual and the technological issues involved in designing and building a research-oriented scholarly digital edition.

The philological motivation for “Make *your* edition: models and methods for digital textual scholarship” is to empower textual scholars to work in computationally competent ways that build upon the foundational merits of such technologies as TEI-XML for authoring and XSLT or XQuery for transformation. But instead of emphasizing specific tools or frameworks or languages, the Institute will train participants to engage with their textual sources in ways that are guided by their individual research goals, using both general-purpose and specialized programming languages to construct the analytic and publication resources they may need to realize research goals not anticipated in existing toolkits. Digital Humanities methodologies that prioritize shared tools and solutions offer a promise of interoperability and interchange that may meet the needs of some users, but learning only to use existing tools comes with a limitation rarely addressed explicitly: editions are motivated by different research agendas, which require different research methods, and the customization or adaptation of shared tools and frameworks needed to meet project requirements fully may wind up competing with and compromising interoperability and interchange. Furthermore, the homogenizing effect of large-scale common solutions can actually discourage innovative thinking, so that the framework may come to determine the direction of the research, instead of letting the scholarly agenda drive the innovation of new technological solutions (Zundert 2012).

The pedagogical motivation for the Institute is *sustainable training*. Tools and programming languages come and go, and training users primarily to use specific tools and frameworks tells them what to do today, but fails to equip them to engage creatively and independently with tomorrow. We italicize “your” in “Make *your* edition” to emphasize that the design of the edition can and should start from the research goals of the individual edition, and that training digital humanists to conceptualize and realize editions built around their own needs, instead of pouring their texts into generic frameworks, is how scholarship, methods, and practice evolve. Our coordinated

focus on theory, method, and praxis is not bound to specific tools and practices; it is a sustainable model that will empower the participants to adapt to inevitable change in the available tools and toolkits.

The Institute draws on our experience in training scholars in (primarily European) programs such as DARIAH, CLARIN, DiXiT, NeDiMAH, and Interedition, in which all of the instructors have been involved. These instructors are leaders in technologies in several key communities and projects, such as TEI (transcription and markup), StemmaWeb (text genealogy), CollateX (collation), eXist-db (XML database and application framework), etc. The instructors are not only innovators and pioneers in scholarly digital technology; they also have decades of experience in disseminating these technologies to humanities scholars in TEI training workshops (TEI XML markup, XSLT transformation of TEI, TEI publishing), in Interedition and other bootcamps and symposia (The Hague, Münster, Sydney), by way of individual and plenary tutorials in DiXiT (The Hague), and in papers and publications (e.g., Haentjens Dekker 2015). The proposed Institute responds to national and regional needs by providing the participants (primarily from institutions across the US; see the discussion of participant selection below) with access to innovative digital textual scholarship theory, method, and practice that has heretofore been less available to US scholars than to Europeans, who have had access to EU programs sponsored by DARIAH, CLARIN, COMSt, DiXiT, Interedition, NeDiMAH, and others.

The impressive success of markup languages, and especially of TEI XML, might seem to suggest that digital scholarly editing is a solved problem. However, in recent years the methodological and theoretical discourse and debate about dynamic, open-ended, and computationally modeled and enabled digital scholarly editions has continued to expand (see, e.g., *Variants* 10, Appollon 2014, and the program of the March 2016 DiXiT convention). As theory and techniques develop, scholars continue to augment the established XML model of hierarchical text (e.g., to represent overlap and discontinuity as a natural part of the data model), and textual analysis and visualization require advanced strategies for exploration and transformation other than, for example, XQuery alone. To be sure, progress in the theory and practice of edition, of the sort described in the publications cited in Appendix I of this narrative, is built on top of these established and productive foundations. However, the ability to envision and implement new types of editions that support new types of research requires from scholars algorithmic knowledge and general programming skills. The tools we use today are unlikely to remain the tools that our students will use for the rest of their careers, and for that reason our focus on envisioning and then building new types of editions motivated by new research questions is the foundation of a robust, adaptable, and sustainable agenda for digital edition. This skill set will allow scholars to express their methods in a context where learning to use a specific toolkit is not enough to prepare for an expanding future (Jones 2014, Berry 2014).

Computational skill sets in Digital Humanities vary widely. Not all participants in an edition project need a full range of technical skills, but if a project is to break new ground in the theory of edition, and not merely pour new text into existing templates, the person on the team with responsibility for determining the goals and purposes of the edition needs to understand fully both

the philology and the technology. The Institute caters to this varied constituency by moving from the basic digital environment to full digital publication of scholarly editions. An optional first-week bootcamp establishes basic skills (files, file systems, sharing resources and code responsibly and collaboratively, etc.) for those with limited prior experience on the command line. The second week allows the participants to practice and build on basic skills when they combine digital textual scholarship theory (e.g., McGann 2004, Siemens 2012, Robinson 2013, Haentjens 2014, Andrews 2012, Pierazzo 2015) with standard (e.g. XML, Python, Jupyter Notebooks) and advanced (e.g. StemmaWeb, CollateX, NEO4J, Tinkerpop, eXist-db) digital technologies. By the end of the third week participants will be able to conceptualize and create an edition that meets their research goals by engaging programmatically and algorithmically with their data and analyzing and visualizing it in ways that transcend, and do not merely enhance, the production of an annotated reading text.

The modularization of the Institute into an introductory basic skills bootcamp and a two-week intensive training course in new digital textual technologies caters to an audience of scholars with mixed prior experience. Novices in the computational side of Digital Humanities are welcome to join for the bootcamp, while those who are more experienced can opt in at the second week. The course will also offer an optional day-long disclosure of the didactic and pedagogical underpinnings and approaches of the course itself. This train-the-trainer approach will allow scholars attending the Institute to re-run personal variants of the course, so that they can share the transferable skills and knowledge they will acquire in the Institute with their colleagues and students at their home institutions.

## **B. Institutional Profile**

The Institute will be held at the University of Pittsburgh main campus, located in a major urban center with an international airport and reliable train and bus service. The University of Pittsburgh (Pitt) is a leading coeducational, state-related public research university founded in 1787. A member of the Association of American Universities, Pitt offers a wide range of academic programs and services for Pittsburgh's metropolitan area population of 2.4 million and its enrollment of nearly 35,000 students. The University comprises 16 undergraduate, graduate, and professional schools, and its 25 libraries and collections have more than 6.6 million volumes. Pitt has moved into the top 10 American higher education institutions in terms of federal funding, as reported by the National Science Foundation.

Institute sessions will be held in a classroom equipped with computer projection and the electrical and network connectivity needed to support all Institute activities. The Institute Director teaches a course at Pitt in computational Digital Humanities, cross-listed in eight departments, and two undergraduate assistants, recruited from graduates of that course, will contribute to the Institute by helping to prepare instructional materials, circulating during hands-on Institute lab sessions, and guiding participants through individual sticking points. The Digital Scholarship Services section of the University Library System (ULS) has agreed to host the materials developed for the Institute, ensuring open access, stable institutional URLs, and a commitment to continued, sustained preservation.

Participants will be housed in the University's Bouquet Gardens dormitory apartments in private rooms with private baths and provided with meal plans in the University dining services. Institute participants will have full access to the University libraries, including networked library and other resources.

### **C. Curriculum and work plan**

The Institute is divided into three one-week modules and is targeted toward participants who have experience with TEI XML or similar textual markup preparation, but who lack the expertise to turn a tagged text into an innovative digital edition. Week 1 is optional, and participants who already have the skills described below can join the Institute at Week 2. Week 2 concentrates on transforming tagged text (such as text in TEI XML) into a richly annotated digital edition suitable for scholarly exploration that exploits the unique capabilities of the digital environment. Week 3 uses the intelligence introduced into the tagged text in Week 2 to publish an edition that supports the intended research.

To illustrate what participants will be able to accomplish after each day of the Institute, Appendix B provides persona studies of two hypothetical participants with different research goals and prior experience. The core Institute staff (Director and instructors) will participate in all instructional activities, with designated lead instructors directing each session and the rest of the team assisting participants in hands-on activities, including work with their own project materials. The entire instructional team will participate in all daily activities described below, with each day coordinated by designated lead instructors, selected on the basis of their prior teaching and training expertise in particular technologies: Tara Andrews (TA), David J. Birnbaum (DJB), Ronald Haentjens Dekker (RHD), Mike Kestemont (MK), Leif-Jöran Olsson (LJO), and Joris van Zundert (JVZ). Instructor biographies are provided in Section F of this narrative, with brief professional resumes in Appendix C.

#### **Week 1: Bootcamp: Computational groundwork skills**

**Rationale:** Perhaps the greatest practical challenge in any hands-on workshop in the Digital Humanities is the variation in background knowledge of the computing environment that the participants bring to the class. We put forward a novel solution to this problem, which we hope will be adopted in workshops throughout the digital cultural heritage sector, by dedicating an optional first week to laying a groundwork of computational skills and knowledge that cannot reliably be assumed of digital humanists.

All participants will have been asked to prepare beforehand by working through an introductory Python programming tutorial on the Codecademy website, which we have used as preparatory material in other workshops we have taught, and for which we will provide support before and during the Institute on a discussion board that we will maintain. The goal is not that the participants will arrive as capable Python programmers, but that they will have had the opportunity to operate at an introductory level within a programming environment. But because web-based courses such as those offered by Codecademy teach programming within a browser, once new programmers have written Python code, how do they run it? Test it? Re-use the logic elsewhere? Indicate that a program should access a particular file on the computer or on the Internet?



Interpret and respond to cryptic error messages? The first week of our Institute will provide intensive hands-on practice within the computing environment, so that all participants will be able to participate fully in the digital scholarship activities that form the focus of the following two weeks.

Participation in the first week of the Institute will be optional. Those who already have this basic technical background knowledge can begin the Institute in the second week without any disadvantage, and they will be joined by participants who will have had the chance to acquire confidence in their ability to keep up with their “power user” peers. We regard this *bootcamp* strategy as an innovative way to maximize the accessibility of the Institute by enabling scholars with different levels of prior expertise to participate equivalently in the digital-editions core activities, to which we dedicate the second and third weeks.

**Day 1: Working with files, programs, and the command line.** Digital humanists with advanced expertise in some areas, such as editing files in an XML editor, often have little or no meaningful understanding of the way files are organized on their computers, or of how to interact with them from the command line. On Day 1 TA and DJB will coordinate the introduction of the participants to the command line and hierarchical file system as part of the working environment in Windows, Mac OS, and Linux.

**Day 2: Regular expressions and operating system conventions.** Regular expressions are an indispensable text-search mechanism used in word processors, text editors, and programming languages, and many of the conventions that arise in regular expressions are influential elsewhere in the operating system (e.g., the use of the “\*” character to mean “everything”). On Day 2 TA and DJB will coordinate the introduction of the participants to using regular expressions on the command line and in programming languages.

**Day 3: Understanding programming and naming conventions.** Anyone who has taught Digital Humanities has encountered students whose markup or code doesn’t work because they’ve written it in Microsoft Word, which has turned their straight apostrophes and quotation marks into curly, typographic ones. Programming requires the use of an appropriate editor and an understanding of how appearances can be deceiving (e.g., a tab that looks like four spaces to a human doesn’t look like four spaces to a computer), and an understanding of how to manage names (of files, functions, variables, etc.) in a way that respects reserved words and special-function metacharacters. On Day 3 TA and DJB will coordinate instruction in formatting, file-naming, and command-line character conventions.

**Day 4: Understanding how the Internet works.** With the growth of HTML5 and Web apps, digital publication on the Internet continues to overtake CD and DVD publication, and digital humanists often interact with those editions using Internet protocols of which they’ve never heard. Users of editions may not need to understand how the Internet works, and textual editors may not require comprehensive knowledge of how networks operate, but developers of digital editions that aim to offer more than a static Web archive *do* need to understand the basics of network connectivity and the typical architecture of a Web application. On Day 5 RHD and JVZ

will coordinate instruction needed about network domains, ports, and protocols to deploy complex web-based digital editions.

**Day 5: Don't Panic: Documentation and error messages.** Error messages originate in a variety of places and the text they produce is not always easy for humanists to understand. Dealing with errors and error messages doesn't require a degree in computer science, but it does require a knowledge of how to read and understand the message. On Day 5 RHD and LJO will coordinate instruction in how find the location and nature of coding mistakes in seemingly inscrutable stack traces and other error reports.

## **Week 2: Philcamp: Digital philology, creating a digital edition**

**Rationale:** Our experience teaching in TEI and other XML workshops has shown that learning nothing more than tagging may leave scholars staring at their angle brackets and wondering what to do next. For some a framework solution like TEI Tapas or TEI Simple provides an adequate next step, but digital humanists cannot build editions that break new methodological ground solely on the basis of template-oriented solutions prepared by others, and because tools and technologies come and go, learning to translate original digital thinking about editions into original implementations of those editions (instead of just “tooling up” on today's applications and platforms) is necessary for *sustainable* training and education.

Recent digital editions training programs (such as the EU DiXiT project, in which several of the instructors have taught) and publications (such as Pierazzo 2015) have shown that digital technologies enable not only new editions, but also new *types of edition*, capable of answering research questions that could not have been asked previously. The goal of Week 2 is to empower the participants to create editions that are envisioned and constructed to meet project-specific research questions, and that are not constrained by predetermined models. The Institute instructors are pioneers in the development of digital editions that move from a grounding in research questions to innovative perspectives on those questions to imagining and implementing new resources to realize new types of editions. The hands-on portion of this Institute introduces tools that we have found valuable, but the goal is not as much to teach those specific tools as to help the participants learn how to move from text and research question to vision to method to implementation—whether that involves the adoption of existing toolkits or the creation of something new.

We divide the second week into five major topics. On each day we will explore the reflections of one of those topics in exemplary existing projects, starting in each instance from the methodological premises (whether implicit or explicit) that underlie specific solutions. Participants will learn how to think about the five topics in the context of their own research questions, and they will gain experience in applying the daily themes to their own materials. By the end of the week the participants will have acquired significant hands-on experience with the tools we use for illustration, but what is more important is that they will have gained experience in conceptualizing, planning, and implementing the transition from tagged text to digital edition within the context of their own project-specific research questions.

**Day 1: Modeling.** On the first day we emphasize three foundational aspects of digital textual scholarship: text as graph (TAG), theory of edition, and the Gothenburg model of textual variance (GM). This day will also cover developing project specifications and defining goals and “non-goals” to avoid “scope-creep”.

*Text as graph:* The general framework we employ for discussing digital approaches to textual scholarship is text as graph (TAG), which allows the user to model arbitrary relationships among pieces of a text. In XML the information has to be stored in a hierarchical context, while with a graph model no such hierarchy is imposed (although the information can be structured hierarchically if one chooses, which means that the graph model is backwards compatible with the XML tree model). We employ TAG in the Institute to model structures that are important in documents whether or not they are isomorphic with an XML tree model. On Day 1 RHD and TA will introduce participants to modeling their own texts as graphs, and to understanding how the graph model enables them to ask questions of those texts.

*Theory of edition:* Textual editing requires attention to the relationship of manuscript to text, to documentary vs analytic functions, to text and language, to exemplar and copy, and to what we have called a reader-oriented theory of edition (Sels and Birnbaum, in press). A scholar should seek not to edit and publish a text, but to edit and publish a text *for a set of reasons*. On Day 1 DJB and LJO will guide the participants through conceptualizing a research-driven, user-oriented edition.

*The Gothenburg model of textual variance (GM):* GM emerged from a 2009 symposium within the frameworks of the EU-funded research projects COST Action 32 and Interedition, the output of which was the modularization of the study of textual variance into stages: tokenization, normalization, alignment, analysis, and visualization. On Day 1 RHD and JVZ introduce GM to the participants. We address all of its parts in greater detail during the rest of Week 2, with continued attention to visualization in Week 3.

**Day 2: Transcription, tokenization, and normalization.** Manuscript materials must be digitized as character data for subsequent computational processing as text. Transcription, tokenization, and normalization are foundational issues in philology independently of computation, but the digital environment entails additional assumptions and consequences. On Day 2 DJB and TA will coordinate the introduction to computer character set conventions (plain vs fancy text, character vs glyph, Unicode, etc.), and RHD, LJO, and JVZ will guide the participants through the division of continuous text into smaller units for subsequent comparison (tokenization), and strategies for letting the research goals of the project guide decisions about normalization. This portion of the Institute will also explore *contextual normalization*, that is, ways of exploiting textual distinctions for some purposes while ignoring them for others.

**Day 3: Collation.** We discuss the alignment stage of GM, or collation, within the data model of the Variant Graph. On Day 3 TA and JVZ will coordinate an introduction to using the open-source CollateX framework to align the participants’ own documents. DJB and RHD will discuss collation as both a philological process (determining what to align with what) and an engineering

challenge (negotiating the ambiguities and complexities raised by duplication, transposition, or order effects, exact vs fuzzy matching, etc.).

**Day 4: Markup and annotation.** The analysis stage is the researcher’s opportunity to interact with the automated alignment. This day will explore ways to build on top of XML structures (for example, using ranges or graphs) to represent properties that are difficult to express and process efficiently solely within the XML tree model. Day 4 will draw attention to layers of annotation (base, linguistic, named entity, etc.), examining how annotation layers interact within an edition. Participants will learn about customized annotation (e.g., through XPath), about natural language processing tools (Stanford, NLTK) that can be recruited to support linguistic annotation, and about the integration of those formerly stand-alone, plain-text applications into an XML environment (such as the recent incorporation of the Stanford NLP tools into the eXist-db XML database by LJO, who, with RHD, will lead this day’s activities).

Beginning in Week 2, Day 4 and continuing through Week 3, Day 1, the Institute instructors will be assisted by MK, a founding member of the Computational Stylistics Group, whose expertise in stylometry during the transition from Week 2 to Week 3 will supplement the digital edition expertise of the regular Institute instructors. On Week 2, Day 4 MK will introduce the participants to annotation-related prerequisites for the statistical exploration and analysis of textual information.

**Day 5: Query and exploration.** Week 2, Day 5 develops a query functionality that lets the user express research questions in a way that addresses relevant annotation layers and retrieves information in a variety of formats (e.g., table, tree, graphic visualization). To the end user this looks like an XML database, but underneath it can integrate tree, range, and graph modeling. On Day 5 LJO and DJB will coordinate the discussion of textual queries and the retrieval of textual results, while MK will focus on statistical queries and the graphic visualization of their results.

### **Week 3: Pubcamp: Expressing the edition, digital publication**

**Rationale:** In Week 3 participants will learn to transform the deeply annotated edition designed and constructed in Week 2 into the enriched digital publication that we characterize as a *digital workstation*. Participants will gain hands-on experience expressing their editions according to W3C Web standards, but the focus will continue to be on theory and method, in this case on learning how to conceptualize the final stages of publishing an edition that meets the scholar’s specific research goals.

Week 3 builds on the querying, filtering, analysis, and transformation of XML and graph data that concludes Week 2, but while in Week 2 the goal was the *developer’s* exploration of the edition, in Week 3 the developer learns to publish an edition that supports querying, filtering, analysis, and transformation *by end-users*. Among other things, participants will experiment with using eXist-db, one of the first implementers of TEI Simple, in the expression and workflow of rich source materials. An important property of digital editions is their potential support for graphic visualizations that expose features of the text that are obscured behind the “wall of words” that constitutes the primary reading view. For this reason, Week 3 conceptualizes the reading view as

just one of many possible views of the text, and one that may be produced alongside others and integrated with them.

Week 3 is divided into five major topics, which we explore in a scaffolded way, leading to a richly integrated understanding of digital publication. Participants will explore the five topics both abstractly and in the context of their own data and research questions, and they will gain experience in applying each day's topic to their own materials. By the end of the week the participants will have acquired significant hands-on experience with expressing their edition both structurally and analytically.

**Day 1: XML and data analytics.** Expressing the edition as XML. Topics include using XQuery to query, filter, and analyze the text as data (cf. Ramsay 2011, Jockers 2013), in preparation for delivery as HTML in Day 2. XQuery has strong transformation capabilities that will be used for this day (to be augmented by XSLT transformation on Week 3, Day 3 for book publication). The querying, filtering, and analysis, coordinated by TA and DJB, will establish a foundation for publishing an API (application programming interface) on Day 4. Participants will gain experience thinking creatively about the graphical visualization of textual information and the results of analytical explorations. MK will contribute to the day's activities by concentrating on the preparation of graphical views of statistical information for publication, and on their integration with text-oriented views of the same underlying data.

**Day 2: HTML and Web components.** This day builds on HTML and CSS from Week 1 and query explorations from Week 2, Day 5. Participants will learn to employ Web Components, a set of standards that allow reusable resources in Web documents and Web applications, thus bringing component-based software to the Web. This day, coordinated by DJB and JVZ, provides experience in using the information queries explored and developed on the previous day to determine uses of Web Components for expressing (that is, publishing) their editions.

**Day 3: Book publishing.** One of the most valuable features of digital text technology is its inherent support for multipurposing, that is, for the use of common input files to generate multiple output views and perspectives in multiple media. On this day, coordinated by LJO and JVZ, participants will gain experience performing single-source publishing of their editions, learning how to generate LaTeX, EPUB, and PDF output to complement the Web editions that form the focus of the preceding and following days. These different formats are treated in a coordinated way by focusing on transformations and styling, so that although the output format and medium will vary, the basic architecture of the process will remain constant. Week 3, Day 3 will also explore single-source publishing with XSLT, XSL-FO, and TEI Simple.

**Day 4: Publish an API.** One of the cornerstones of innovative digital publication is what we have termed a *reader-oriented theory of edition*, where the edition supports multiple methods for navigation and exploration and enables the reader to participate in shaping the reading experience. Building on previous topics and on the general scaffolded process that informs the design of the Institute, we introduce the participants to consolidating their queries and the results of their analysis into an API, which will guide the eventual reader's interaction with the edition. This day, coordinated by TA and RHD, will explore the theory underlying the design of an effective

API, one that will be capable of expressing the edition's specific information and visualization needs. This day will also discuss practical issues that need to be addressed and resolved before publication, including licensing, archiving, and end-of-life maintenance.

**Day 5: Hosting and containers.** Digital editions must be made available to users, but many academic institutions are unfamiliar with hosting anything other than static Web pages, and many digital humanists lack meaningful experience with the stewardship, preservation, and maintenance issues that determine the accessibility and sustainability of the edition, and thus its scholarly impact. Topics to be covered include Infrastructure as a service (IaaS), Platform as a service (PaaS), Software as a service (SaaS), containers, and cloud hosting. In this session, coordinated by RHD and LJO, participants will learn how to make hosting decisions based on the needs of the edition.

#### **D. Participant selection**

Participants in this Institute will be recruited via an open call on DH mailing lists including Humanist, Digital Classicist, Digital Medievalist, TEI-L, WWP-Encoding, and others. A draft *Call for applications* is included in Appendix F.

The target audience is scholars who are acquainted with some aspects of digital textual scholarship (most commonly by editing documents using TEI XML), but lack the knowledge of how to transform a tagged text into a published digital edition. Because digital edition training and instruction is more accessible in Europe (through EU initiatives sponsored by DARIAH, CLARIN, and others) and Canada (especially through the DHSI) than in the US, and also for budgetary (airfare) and logistic (visa) reasons, we expect that most participants will be US-based scholars, although to the extent that the budget will permit we will also give full consideration to highly meritorious non-US applications.

Applicants will be asked to describe a digital edition project that they are involved in or want to undertake. We are especially interested in proposals from scholars whose projects are driven by innovative theoretical research questions, and whose needs therefore may not be met by the simpler solutions offered by frameworks like TEI Tapas or TEI Simple. We will, therefore, expect applicants to have both a willingness to acquire computational skills and the ability to identify and acquire data sets for their projects that are amenable to computational analysis and visualization. Recognizing that many digital humanists who are comfortable tagging texts in TEI XML may be unfamiliar with other basic features of the digital editing environment (e.g., the command line, the hierarchical file system, etc.), we will invite participants to apply for either the full three-week Institute or, in the case of those who have already acquired the basic skills to be emphasized in Week 1, only the second and third weeks. We anticipate accepting 15–20 applicants for the full three weeks and an additional 5 for just the second and third weeks.

We will aim for a personally and professionally diverse participant group, and we will follow the policy adopted by other NEH-funded Institutes and reserve four to five spots for advanced graduate students in the humanities. The methods and tools that we will teach in the Institute will help them advance their research at the very beginning of their careers, while also providing them

with a practical skill set that will augment their traditional academic credentials in the job market when they complete their studies.

The *Call for applications* will be disseminated on January 1, 2017 with an application deadline of February 28. Selections will be made by March 15 after review and discussion by the five Institute instructors. The selected participants will be asked to install software, read materials, and complete the free Introductory Python programming tutorial on the Codecademy website, all of which we have required as preparatory activities in other workshops we have taught, and for which we will provide support before the Institute on a discussion board that we will maintain. The expectation is not that the participants will arrive with developed Python programming skills, but that they will have gained enough preliminary experience that we will be able to concentrate with them on the features that will be important specifically for the digital edition goals of the Institute.

### **E. Evaluation and impact**

The expected outcomes of the Institute include the following:

- **Participant outcomes.** Participants will learn to conceptualize and implement not only new editions of new materials, but new *types* of editions that enable new types of research questions. It would not be realistic to expect participants to become expert in all of the Institute content in three weeks, but our extensive teaching experience has confirmed that it is fully realistic to expect them to learn enough and gain sufficient hands-on experience to undertake new types of editions, and to know where and how to look for more information when needed.
- **Impact on Digital Humanities scholarship and practice.** Innovative digital edition theory, method, and praxis is more widely available in Europe than in the US because EU funding models have been very strongly supportive of it. The Institute instructors, leaders in the European digital edition community, have an established research and teaching relationship with the Project Director that will enable them to collaborate to narrow the gap between US and EU digital edition scholarship opportunities. The Institute thus meets an important national need by making available to US participants a training program that is otherwise far more accessible in Europe.
- **Impact on Digital Humanities pedagogy.** We expect the Institute to advance Digital Humanities pedagogy in the following ways:
  - **Digital humanists can learn computation and programming.** Our experience teaching content similar to that of the Institute has confirmed that the degree of computational and programming knowledge needed to empower digital humanists to imagine and implement their editions without ignorance or fear of technology beyond markup and canned applications is not beyond their reach. Our Institute implements a Digital Humanities pedagogy that includes “if you can think originally about what you want to accomplish with your edition, you can learn to build it.” In other words,

the research question must precede decisions about tools, lest a dependence on familiar tools come to discourage new directions for research.

- **Task-driven learning can yield high-quality results quickly.** Teaching a programming language or tool thoroughly brings learners to the broadest level of practical proficiency, but much of what digital humanists need to do can be accomplished with expertise that is narrow but deep. Digital Humanities pedagogy can cover a large number of technologies in a short amount of time by recognizing that it's possible to learn how to do something difficult and specific without having to learn the technology comprehensively. Part of this training involves learning how to learn: how to break down a task into components, look things up, debug, etc. We expect that our Institute will serve as a model for teaching digital humanists to complete sophisticated tasks without a degree in computer science, or even a comprehensive knowledge of a programming language (Birnbaum 2014).
- **Institute participants can become teachers.** One innovative aspect of our Institute is an optional day-long pedagogical recapitulation of the course content. Participants in previous workshops we have taught have gone on to become Digital Humanities teachers themselves, and this “train-the-trainer” coda to our Institute will prepare participants to share with others what they will have learned from us. In this they will be aided by the sustained free availability of our teaching materials, as described below.
- **Sustainability and accessibility of materials.** All materials produced for the Institute will be published under a free cultural works Creative Commons Attribution Share-Alike license, which makes them available at no cost to all interested persons. The University of Pittsburgh Library System (ULS) has made a commitment to maintain all teaching materials after the conclusion of the Institute as a Web Archive. These resources will have stable institutional URLs and a commitment to long-term maintenance and will remain available for free use in future Institutes, courses, and workshops, and as tutorials for individual consumption.
- **Assessment.** Assessment of the effectiveness and impact of the Institute will include the following models:
  - Each day's activities will conclude with a brief online self-assessment by the participants of the extent to which they have achieved the learning outcome goals of the day (a sample assessment form is included in Appendix G). This formative evaluation will enable the instructors to gauge the effectiveness of specific instructional methods and materials.
  - The most meaningful measure of the effectiveness of the Institute within the digital textual scholarship community will be the extent to which the participants use the Institute experience to conceptualize and implement their own innovative digital edition publications. One year after the conclusion of the Institute we will ask the participants



for a report on the influence of the Institute on their digital edition methods and practice. This summative information will help us assess the more durable impact of the Institute.

- **White paper.** The Institute will produce a white paper for publication on our Web Archive and on the NEH website. The white paper will document the Institute, including lessons learned, so that others can benefit from our experience.

#### **F. Project faculty, staff, and consultants**

The Institute Director, **David J. Birnbaum**, is Professor and Chair of Slavic Languages and Literatures at the University of Pittsburgh, where he teaches a course on “Computational methods in the humanities.” David is responsible for the overall design and organization of the Institute, for coordinating the activities of the faculty, staff, consultants, and participants, and for contributing to the instruction. David has been actively engaged in Digital Humanities scholarship and teaching for almost 30 years, including two two-year elected terms on the TEI Council, and he is responsible for the design and implementation of open-access digital editions of the *Rus' Primary Chronicle* and the Old Church Slavonic *Codex Suprasliensis*, a UNESCO Memory of the World monument.

For the past several years David has collaborated in research and teaching with the prominent European digital philologists who constitute the rest of the instructional staff of the Institute:

**Tara L. Andrews** is Assistant Professor of Digital Humanities at the University of Bern, where she teaches a series of courses that introduce students to digital tools and methodology. She obtained her D.Phil. in Oriental Studies at the University of Oxford in 2009; she also holds an M.Phil. in Byzantine Studies (2005) from Oxford and a B.Sc. in Humanities and Engineering (1999) from the Massachusetts Institute of Technology. Tara previously worked at KU Leuven on the *Tree of Texts* project, an investigation of the theory behind stemmatic analysis of classical and medieval manuscript texts, and she maintains the freely available suite of tools developed for the project as *Stemmaweb*.

**Ronald Haentjens Dekker** is Head of Research and Development at the Huygens Institute for the History of The Netherlands, part of the Royal Netherlands Academy of Arts and Sciences. As a software architect, Ronald translates research questions into technology or algorithms and explains to researchers and management how their choice of technology will influence their research in the future. Ronald has worked on transcription and annotation software, repository software, and collation software, and he is the lead developer of the CollateX collation tool. He also conducts workshops and training for researchers in computational thinking and the use of scripting languages in combination with digital editions.

**Leif-Jöran Olsson** has been employed since 2005 as a systems developer at Språkbanken, the Swedish Language bank, University of Gothenburg, where he develops research infrastructure for language technology, both nationally and within CLARIN ERIC. His project management experience involves both long-term partner projects (e.g., the Swedish Literary Bank, the Selma Lagerlöf Archive, the Swedish Drama web) and short-term domain-specific toolboxes (including training and use case analysis). Leif obtained his MA in Language Technology from Uppsala

University in 2004. He is one of the core developers of the open-source eXist-db native XML database.

**Joris J. van Zundert** is a researcher and developer in the field of digital and computational humanities at the Huygens Institute, where he serves as Program Leader of the Methodology Research Program, which fosters discussion and research on textual scholarship, documentary editing, and their relation to digital and computational methods. Joris's main research interests focus on the role of computational algorithms in the analysis of literary and historic texts, on the nature and properties of humanities information and data modeling, on computer science and humanities interaction, and on the tension between hermeneutics and "big data" approaches.

The core instructional team will be joined at the end of Week 2 and the beginning of Week 3 by **Mike Kestemont**, Assistant Professor in the Department of Literature at the University of Antwerp in Belgium. Mike's research focuses on computational text analysis of historic texts, with particular attention to medieval texts from western Europe, and he has taught courses and workshops on digital text analysis, corpus and computational linguistics, programming for the humanities, and medieval philology. Mike is a member of the Computational Stylistics Group and a developer of the open-source Stylo (Stylometry in R) toolkit and the PyStyl Python package for Stylometry.

## Appendix IV: July-September 2018 participant evaluations of 2017 IATDH

As described in the body of the white paper, one year after the conclusion of our Institute we wrote to all participants to request feedback about how their experiences at the IATDH had contributed to their subsequent research and teaching. The full text of the twelve responses that we received is below.

### Susanna Alles Torrent

*What was your general experience as a learner during the Institute? If you participated in the bootcamp, we would be particularly interested in any specific comments you might have about it.* My general experience during the Institute was outstanding. It gave me a quite complete overview of the processes and the skills needed to implement a digital edition. I really enjoyed the step-by-step organization of the Institute and the care that the instructors gave to make you understand the different phases that a digital editor has to overpass. I participated to the Bootcamp during the first week and it was a great refreshing start. Basic concepts and the possibilities of the operating system, the utility of the command line and the regular expressions and its potential to manage (text) files was then clear to all students. Also, becoming familiar with Git and GitHub was essential, and it was a great decision to use it in the classroom, allowing an optimal organization both for instructors (that could update their materials), and for students, always ready to download the latest version of the materials and being able to personalize their own

copy. The Python section was for me the hardest, maybe because, even if I have followed several workshop and tutorials, I have not use it in a real project. As any language, you get to know it when you use it and face real life challenges.

*How have the digital edition concepts and skills introduced in the Institute influenced the way you conceptualize, design, and implement your own editions?*

For me, the Institute helped me to better conceptualize and realize the weakness of my project, as well as to picture the broad array of the available tools, that I could implement in my own edition. Listen to the instructor and students' experiences help me to identify many challenges that we all share but also new ones. Many sessions, such as the one devoted to collation, gave me helpful ideas that I now consider essential and a must to be included in my project. I really appreciate the introduction of initiatives that I was unaware of, such as the CETEIcean or Flask. In the tokenization part, I would have added a lemmatization section, I think that NPL contents would improve the already excellent content of the Institute.

One of the parts with which I felt overwhelmed were the eXist database sessions. I understood the potential of the software, and I got a general overview of the process. The first challenge was to understand XQuery (which maybe should be introduce in the Bootcamp?), and it would have been useful to work further with one example given by the instructors and work a little bit more on the final project, that is, how do I implement eXist on my web structure and how do I make it work on the internet and in my webpage-digital edition published online.

*As we wrote last fall, we are maintaining a page of links to digital edition projects (whether completed or in progress) by members of the Institute at [https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant\\_editions.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant_editions.html), and we would like to include information from as many participants as possible. We would be grateful for information about any edition projects you are pursuing, and about how your work at the Institute has contributed to those projects. And if your edition (whether complete or in progress) has a web presence, please let me know and I'll be happy to link to it (and to update any references that are already there).*

I have been working for a while on the conceptualization and the implementation of a digital edition for a set of biographies written in Latin and translated in old Spanish during the XVth century. The challenges that I had were, among others, to face a text in translation, offering the original and a version, and explore the possibilities of linking both texts; and, also to consider the best way to create a digital infrastructure to publish the digital edition. You can find the first attempts of this project at <https://alfonsodepalencia.github.io/Vitae/> Hopefully, and when time permits, I will change the current architecture of the site, adopting an eXist db and a search engine, adding features such as the collation and visualization of the Latin manuscripts collated in this edition, a better design, and a tokenization and lemmatization feature.

*How have you used the skills and materials from the Institute in your own teaching, either in the classroom or in collaborative projects? If you have developed a course description or project guide that was influenced by the Institute, we would be grateful if you would share it with us.*

In my case, it was particularly useful the first and the last part of the Institute. In my teaching, it is extremely valuable to have the online materials still available as far as the general sections (how the internet works, and Program and files), and especially the command line and the regular expressions are concerned. Then, and when I teach TEI, I have decided to include a section for the Web publishing frameworks, such as CETOcean, GitHub Pages, or Flask.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

As I have mentioned, intellectually I feel more aware of the many steps that one has to undertake to conduct a digital edition, and the skills needed to implement (it is not enough to know TEI, but there is a full array of other technical skills that benefit the conceptualization of the digital edition). Professionally, I am more confident on the tools that we have at our hand to implement ideas and the real process and the dimension of the labor that we have to face.

*We would be grateful for any feedback you might be able to share about the scheduling and logistics (both academic and practical) of the Institute.*

The Institute was very ambitious in scope, but I think that the scheduling and logistics were quite excellent. Maybe it would be interesting to have a real example running through all the Institute, trying to connect all sessions with one example. For instance, how python has been used in the Sample Edition, how the Semantic Web feature has been implemented, why an API is valuable, how the eXist db is enabling the web infrastructure and the search engine, how the tokenization is used, etc.

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

As said, the Institute was a great experience and it did accomplish one of its main goals which was "learning to think and act digitally about the process of creating a digital edition." In such short time period, students got a tremendous amount of information, both from a theoretical and practical (software, tools, languages, etc.) point of view. For this reason, I would suggest that the Institute was based in a single and concrete example, maybe built ad hoc, that somehow would contain all technologies learned, in order to always have a real and handy example. In the case of some frameworks, which have required much time and practice, as the eXist db, would have been useful to insist in a real implementation and arrive to the final step of the web publishing. Finally, I must say that I really appreciate the organization of the Institute under the form of a website, and especially that we can still access to all the instructor's materials. I use it quite often for my work.

## Halila Bayramova

Please, allow me to use this opportunity to thank you for keeping in touch and providing a chance to reflect on one's work a year after NEH Institute'17. This period coincides with the second year of my studies on the PhD register, working towards a digital genetic edition of James Joyce's *Finnegans Wake* Chapter II.2.1 (For reasons of conformity and, to a lesser degree, academic elegance, it has been decided to refer to the 10th chapter of *Finnegans Wake* as "Chapter

II.2”. I apologize if my inconsistency caused any confusion.) The end of this academic year marks an official submission of the first thesis chapter (as part of the requirements for a successful confirmation (off-the-record: a Trinity exam for secondyear postgrad students, aka “mini-viva”)). Also, a draft version of chapter two has been presented at the Antwerp James Joyce Symposium in June 2018.

This year has been much about disillusionments: if, at the beginning, I was more aware of theoretical frameworks, working within the French tradition of *critique génétique*, within the Joyce editorial history and textual scholarship, this year the project has been bouncing off the copyright restrictions and institutional requirements. But, it is a good kind of disillusionment demanding more resilience, flexibility, and a bit of humour.

The question of how NEH Institute ‘17 has helped to shape my ideas and concepts of a digital edition is analogous to the chicken-or-egg dilemma. As I attended the school right at the start of my project, its intellectual impact has blended in too well and now it is hard to tell which bits and pieces have emerged as a direct result of it. (Probably, all of it!) Therefore, in hindsight, I am glad the organisers asked to draw a preliminary report in Sep 2017, while the impressions were still fresh and unadulterated. The following is my feeble attempt to add to the points mentioned in the first report.

As a newbie with limited hands-on experience, I am afraid I have little to contribute to the discussion of the Institute content. For me, it was a one of-its-kind crash course delivering exactly what its title promised from the onset: “Make Your Edition.” It took me from ground zero to the very last step of digital editing process, mapping my detached chunks of skills and knowledge along the way, creating a logically consistent network, and, providing resources for further self-study to cover gaps in between. I must highlight another important point: it gave me the confidence to discuss my editorial project with a deeper understanding of the correlation between its technical details and their theoretical implications, which until then had only had an intuitive presence.

*-How have the digital edition concepts and skills introduced in the Institute influenced the way I conceptualize, design, and implement my own edition?*

Emerging from my daily interactions, there seems to be a mistaken conception that the genetic edition tries to explain or annotate the text of *Finnegans Wake*, when its main purpose is to enhance the reading experience not to simplify it. As much as *Finnegans Wake* is difficult to read, it was difficult to write too. Joyce wouldn’t have spent seventeen of his intellectually-peak years drawing an elaborate hoax. (Or would he?) The enormous corpus of textual evidence is a strong argument answering this question. Undoubtedly, reading through the genesis of the text opens a completely new dimension providing a better insight into *Finnegans Wake*. Yet, by presenting an editorial model that best represents the genesis of the Wake, I would like to keep it free from any exegesis that could run infinite and cause confusion on screen. To this end, during the Institute’17, I found it useful to see or rather experience many different editorial solutions, models and tools, which helped me to realize what exactly I wanted my edition to do, and (equally important) what not to do. Since there is such a strong emphasis on the infinite possibilities of a

digital edition in the existing critical literature, the Institute's wide-range and all-inclusive scope was a good filter to deepen my understanding of the limits of an edition and cultivate a more eclectic approach. This, together with other institutional requirements (including the ones of the Trinity School of English), have caused a shift in the way the project was originally conceptualized. As a result, my edition's model has become much more modular in order to be more ambitious: a consideration has been given to the whole compositional history of *Finnegans Wake* as opposed to only one chapter; i.e. the idiosyncrasies of Chapter II.2 should not be an impediment in superimposing this model onto the genesis of the other parts of *Finnegans Wake*. (Easier said than done!) This has considerably complicated the task and may have flattened out certain discrete features of II.2. However, hopefully, in the long run, this approach will prove to be more fruitful, if we ever plan to finish the genetic edition of the Wake. The cumulative wisdom of editorial scholarship is such that there is no univocal way to approach construction of a digital edition. And recent studies of many modernist writers, including Joyce, tip the scales in favour of not having one. Every writer's compositional pattern is unique; and the aim is to try to preserve and reflect this idiosyncrasy while simultaneously contending for longevity and interoperability of their digital editions. However, as someone with a philological background, I fell victim to the trend well described by Tara L. Andrews in Variants 10 :

*The method of production, rather than the published form that the resulting editions take, is the practice wherein lies most of the promised revolution within textual scholarship, but it has attracted considerably less attention than the question of digital publication.*  
(2013, 63-64)

And yes, I would process text manually: Python and bash were considered exclusively for digital publication. Regular expressions were not on the horizon yet. And many other simple operations looked downright daunting. So, the Institute has offered not just a set of skills but a completely different way of thinking about text and its analysis. However, this year has also demonstrated that it is not easy to maintain, let alone develop, the newly-acquired knowledge once confronted with mainstream institutional practices and demands. The actual work on a digital edition sometimes feels more like a side-project, overshadowed by theoretical and analytical essays, where I still haven't figured out the right balance of technical language for a predominantly Joycean audience.

Mapping the present Joyce editing:

Before, during, and after the Institute, I have always been painfully conscious of the limits of my work as a one-person project. Whereas there are inevitable institutional reasons for that, the thought of collaboration has been a hovering presence and Institute'17 has emphasised the importance and general goodness of collaborative efforts, not necessarily through its content, but also via the wide range of international projects presented by the participants.

This awareness has pulled me out of the research vacuum and drawn my attention to other Joyce editions in-progress. At the moment, there are three TEI-conformant projects concentrating on manuscript sources.

1. The James Joyce Library under the auspices of the Antwerp University Manuscripts Centre. Referred by its editors as a relational database rather than an edition, it primarily focuses on the

exogenesis of *Finnegans Wake* and is designed “after Joyce’s own writing habits” incorporating four main stages in Joyce’s composition process: sources, notebooks, manuscripts, and published texts (De Keyser et al. 2017, 110). It has an online presence but for copyright reasons is still closed to the public (<https://joyce.uantwerpen.be/>).

2. Another project is the reimagining of Hans Walter Gabler’s famous 1984 *Ulysses* synoptic and critical edition by Ronan Crawley and Joshua Schaüble, the beta version of which can be found here (<https://www.ulysses.pizza/>).
3. Mikio Fuse’s *The Finnegans Wake* Genetic Research Archive (2012-) is a TEI XML-conformant database in progress: a second life for Fuse’s stylesheets built and expanded over the many years of genetic Joyce research.

While, all of these are still in-development, the links between them cannot be ignored. The enormity of Joyce’s textual heritage evidently requires cooperation on many levels. The existence of these projects is encouraging in many ways as an expansion as well as a backdrop to my own work. And employing similar approaches guarantees a rapprochement of scholarship.

In stark contrast, the official debut of a new project during Joyce Symposium’18 has served as a good demonstration of a thorough research coupled with unwillingness to collaborate: James Joyce Digital Archive (<http://www.jjda.ie/>). As the new website holds the results of forty-years of scrupulous and painstaking work of two Joyce scholars (Danis Rose and John O’Hanlon), the limitations of what one can do with the material are heart-breaking. However, the peculiar nature of the editors’ relationship with their colleagues, which was confirmed by the circumstances in which the website was unveiled, makes it difficult to expect any collaboration.

#### Summary

To recap, the following are the points I found important for my own research during Institute’17:

1. Comprehensive approach
2. Critical thinking
3. Importance of the process over the product
4. Provision of self-study resources
5. Eclectic and modular approach to editing process
6. Collaboration

In the age of “full information display,” when almost any conceivable editorial scenario is technically fathomable, the discussion of the tension between affordances and limits of DSE may be enhanced by a critical assessment of their importance. The way DSE are built and mediated to the reader is not necessarily transparent or self-explanatory. Critical thinking reflected at all stages of digital editorial production calls for an explicit declaration of their theoretical implications. Such critical awareness of the digital text may encourage more academic and non/academic engagement, as well as more institutional trust in digital scholarly practices, since, above all, these editions’ ultimate goal is to be of use to a wider audience of readers.

## Elisa Beshero-Bondar

*How have you used the skills and materials from the Institute in your own teaching, either in the classroom or in collaborative projects? If you have developed a course description or project guide that was influenced by the Institute, we would be grateful if you would share it with us.*

The Institute perhaps most strongly influenced my collaboration on the Frankenstein Variorum project, where I've had to do a lot of instruction of my colleagues on the fly on various technologies. I've applied and experimented extensively in the past year with normalizing algorithms for collation thanks to the Institute, and I've had to explain and document the work for colleagues.

Here is a sampling of such documentation (mostly still under development): [https://github.com/PghFrankenstein/Pittsburgh\\_Frankenstein/blob/master/collateX-Prep/Readme\\_SGA\\_CollationPrep.md](https://github.com/PghFrankenstein/Pittsburgh_Frankenstein/blob/master/collateX-Prep/Readme_SGA_CollationPrep.md)

The Institute didn't really affect my university semester teaching very much because I was already teaching people how to develop digital editions using XML technologies, building strongly on prior work with David. In the past summer, however, I collaborated with David on a one-week course at the DHSI titled "XPath for Document Archaeology and Project Management" designed for people conceptualizing or actively working on digital edition projects. A few of the NEH Institute participants enrolled and they have indicated they found the experience helpful for their continuing project development. The course materials are

here: <https://ebeshero.github.io/UpTransformation/> and in many ways, they augment aspects of last summer's Institute instruction.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

I benefited most from the activities prompting us to conceptualize an API interface for our edition projects. For me, this was extremely helpful because it expanded the usual view of a digital edition to help us imagine how best to optimize multiple gateways and interfaces for an edition project. A scholarly edition (whether print or digital) is not just the pages of documents reproduced on screen, but a search interface perhaps most frequently accessed from the index first. Mapping out what we want to do with our APIs was by far the most constructive (and most-needed) activity for me. Learning the nuts and bolts of how to do it is something I was already working on prior to the Institute in my work with eXist-db, but my view of what was possible was expanded by consulting the expertise of the instructors.

Another direct professional benefit for me is an extended collaboration of a group of us Institute participants working on nineteenth-century editions. We are discussing an extended collaboration on linking prosopography data and sharing methods and resources. We've just begun this discussion this summer, and look forward to applying for grant funding to support a face-to-face summit to meet and plan strategies. This may lead to conference and publication opportunities in addition to enhancing linkages among related data accumulated in our projects.

*We would be grateful for any feedback you might be able to share about the scheduling and logistics (both academic and practical) of the Institute.*



The Institute instructors worked diligently on their instructional plans and on revising materials on the fly so we could spend time on specific areas of interest or need. I particularly admire the instructor team's concern and attention to us individually. We all learned a lot more than we might have imagined about how GitHub worked, for example, when Ronald walked us through forks and branches slowly and clearly, and I think several members of the Institute were inspired to try technologies they may have feared or not understood well before. All of us who participated in the Command Line Bootcamp (the optional week 1) benefited immensely from the experience, including me (and I came in with some command line experience). I learned how to write shell scripts and write pipelines and Tara's and Na Rae's carefully scripted hands-on activities here worked wonderfully well.

I do wish that the Institute had discussed XML technologies for document modeling and schema control of digital editions more than it did. Claiming as this Institute did in its rationale (<https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/>) that transcription, tagging, and markup is already being taught and therefore doesn't need to be covered oversimplifies what people need to understand about the document trees they are constructing, as that relates to decisions they need to make in annotating information, organizing metadata, and designing interfaces for reading and data visualization. Concentrate on XPath and XSLT in order to help us envision our editions and their possibilities. I realize this is difficult to do in even a three-week Institute, but this Institute has a unique opportunity, if repeated, to give its participants a strong foundation in the artful manipulation of markup. The lack of discussion of declarative programming methods (XSLT, a powerful method for people modeling documents in markup that is not taught frequently in the Americas) may indicate a problem in what the Institute organizers decided was appropriate to teach in this first round, but I hope to see it in the future because I know the instructors (especially David) are uniquely capable of teaching it and inspiring others to do more with it than they may have imagined. The XPath unit in the Institute was minimal and several participants needed more, so I organized some sessions with various groups after hours to cover XPath axes, predicates, functions--which was still minimal, alas. There should have been more time for this.

Reviewing what we did discuss of markup in week 1, I'm reminded that we concentrated on layers of information and the problems (and the possibilities) of overlapping hierarchies in XML. It may help to reconsider the length of time allotted to this aspect of the course and perhaps turn it to begin a discussion and some demonstration of how to process marked-up data. Because inevitably the Institute participants do not all have good experience with XSLT, showing us how easy it can be to extract a table of marked data, and also how to build a representative HTML page from the same material would be helpful, with emphasis on what is possible and what it might help the Institute participants to build.

Some questions that might seem too basic but perhaps ought to have some attention: How do people conceptualize the organization of documents and metadata in an edition? How do people develop algorithms to navigate and control multiple documents and their interdependencies in an edition project? In scheduling time for topics, consider devoting less time to edge cases and more

time to the work that most of us do in compiling information and documents when we build editions. Perhaps having everyone develop a model of representing two or three witnesses on a simple poem, and reflecting on how to convert data from variants into an interesting and informative visual summary in the form of charts and graphs might be an illuminating exercise.

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

I benefited immensely from the command-line bootcamp and the units on collation and designing an API.

I'd like to see a unit on working with IIIF standards and issues with serving images alongside documents.

There absolutely needs to be a longer unit on XPath and some hands-on coverage of XSLT. To add new things, other things need to be cut. Stylometry was much appreciated and we learned a great deal, but since stylometric methods typically don't depend on the structure of digital editions, this may seem peripheral to the work of \*making editions\* in retrospect. Similarly, the portions of the course taught with eXist-db might have been peripheral because they depend on a specific tool rather than a skillset and ability to survey available tools for serving up an edition or an API.

## Elli Bleeker

### General experience

I attended the full three weeks of the Institute and still the time period felt too short to really delve deep into the topic of digital scholarly editing. This was by no means a defect of the Institute's program: the instructors managed to design a schedule that covered every aspect of digital edition making. The trouble is that in order to do the topic justice, no training program under three years really suffices (and even then). In light of this, I appreciate that the Institute did not pretend to be able to treat every possible component of digital editions in great detail.

Instead they set out to familiarize us, the participants, with a certain way of thinking: a "computational mindset", if you will, which is highly valuable for any kind of research yet not something one usually learns from your average humanities program. For me, the true value of this Institute is twofold: first, acquiring a computational way of thinking and secondly, learning how to apply that way of thinking to scholarly editing. Most of us were already versed in edition making (albeit not always \*digital\* edition making) and familiar with many general editorial concepts. I also considered myself to be an expert in the field of modern textual scholarship, but my research gained more depth and became more inclusive during and after the Institute (see below). As a Digital Humanist, I was already aware of the advantage of using computational methods for humanities research. Yet these methods often focused on completing some specific task or goal (e.g., making a transcription, authorship attribution, collating texts). The Institute, conversely, proposed to reconsider the entire workflow of digital editing. They used the term "computational pipeline" to teach us to think about making a digital edition as a series of independent steps or

stages in a pipeline. Each step represents a specific task: e.g., transcription, tokenisation, normalisation, etc. These separate stages can be strung together to form a pipeline. Each stage also had an specific objective and the instructors showed a number of tools and technologies to carry out these objectives.

The foundation for this mode of thinking still requires a certain knowledge of how a computer works (ranging from operating systems to programming languages to web technologies). This understanding was established during the bootcamp week. During my application for the Institute I was hesitant whether or not I should attend the bootcamp week, because I could already write some Python code and I could more or less work on the command line. I am glad I did attend the bootcamp week: I wouldn't have wanted to miss out the thorough and patient explanations from the Instructors. I realised once more that there's a big difference between being able to work with computational technologies and actually grasping what they *\*do\** to your data. The computational ground skills we learned during those first days are indispensable for anyone who wants to apply computational methods to their research. If you think about it, it makes sense that researchers who use a computer actually understand how their machine operates: software contains so much implicit assumptions that can influence your data and your results significantly. Also, there is always something to learn: I noticed that some of the participants who skipped the bootcamp week because they considered themselves to be adequately versed in computational skills did have significant more trouble keeping up during the following two weeks.

The Instructors emphasised that the goal of the Institute was not to have us work with one specific tool, but rather to have us think in a computational way about editorial work, asking questions like: *"I want to tokenise my text. How do I want to tokenise it? What technology or tool(s) can I use to do so, and how will that technology process my text \*exactly\*? Can I adjust the technology so that it does exactly what I want? How can I inform other scholars of the technology I used, and justify what it did to my text?"*

This way of thinking made me more *\*aware\** of the various transformative processes that underlie a digital edition. Speaking for myself, I became more critical and more conscious of the technologies we use in the editorial workflow. What is more, keeping in mind the concept of a computational pipeline helped me process the various topics treated at the Institute. Sometimes we had too little time to really explore a certain technology, but I could still locate it somewhere in this computational pipeline, which helped me retain the bigger picture. The only (constructive) remark I may have for the instructors was that this educational intention may have been clear to me, but it wasn't for all participants. Some of them really struggled to keep up with a certain technology and were frustrated that they didn't "got it" right away. Certain technologies, like XSLT, take a while to learn, and it could have been stressed more at the outset of the Institute that we weren't expected to master them during the three weeks but rather to *\*understand\** the concept.

**Influence of concepts and skills introduced in the Institute in how I conceptualize, design and implement my own edition(s)**

As set out above, the most valuable aspect of the Institute for me was the concept of a computational pipeline. It genuinely helped me to reconsider my approach to scholarly (humanistic) research. Designing a pipeline is a useful exercise at the outset of a research project because it compels me, first, to think about what I actually want to achieve and then, working my way back, to think about what (technologies/tools) I need to get there. Furthermore, I find this modular approach a "cleaner" way of doing research: splitting up the process into increasingly smaller "chunks" with specific tasks, knowing what tools I could use to accomplish each task, and understanding their input and output helps me to a better grasp my own research in full. It also contributes to a better communication with software developers, because I can now clearly describe what I want instead of leaving it to them to devise based on an abstract description.

Describing my research as a pipeline resulted in a more structured editorial rationale, because I can argue why I made certain technical choices and communicate to my peers (and potential users) how my results came to be. Finally, I believe (hope) that this clear rationale my editorial processes and the subsequent data transformations within the pipeline makes the data of my edition more interoperable and reusable for other scholars. Incidentally, those last two points are largely informed by the instructors' continuous insistence on code hygiene and "documentation, documentation, documentation". I appreciate that they made us think about the (possible) future uses of our data, and that they made us aware of the norms and standards of the software development community (e.g., open source, documentation, clean code, etc). These issues are also relevant for humanistic research, but have not yet been translated to standards for the digital editing community.

Currently, my research largely concerns modeling and not so much digital editing. Nevertheless I have been able to use what I learned during week 2 of the Institute (Philcamp) because one of the topics in week 2 was, effectively, data models. What I took away from this topic was the benefits of considering alternative data models to XML. Not so much because the instructors wanted us to move beyond XML, but because the instructors treated data models the same as tools and technologies: be aware of what they do to your data and you will be able to employ them to their full potential for realising your research objectives.

I learned that sometimes data models and their boundaries can have a strong influence on how you regard your text, and that it pays off to look beyond the limits of a familiar model. In some cases, these limits have become so familiar to us that we tend to overlook the textual features not accounted for in the model.

On the other hand, by reflecting on the ubiquity of XML I also learned that a data model is more than just a technical matter: the support for XML (in terms of tutorials, guidelines and tools) is unrivalled. For alternative data models to be successful, they would need to match XML's community as well.

### **Use of skills and materials in own teaching and/or collaborative projects**

I have and will continue to make use of the concept of a computational pipeline in workshops and teaching. It is a clear and versatile manner of introducing students to a wide variety of digital

editorial concepts while at the same time providing them with an useful abstract perspective on the editorial process.

Furthermore, the deeper introduction into the workflow of git and GitHub has proved itself to be very useful for managing collaborative work. I do think it takes a lot of time before someone is able to employ git to its full potential and there's undeniable value in learning to deal with something like merge conflicts.

This brings me to a third skill that I learned at the Institute: how to solve problems and where to find more information or tutorials for self-study. The instructors made it clear that appearances may deceive for they, too, are not all-knowing and still need to look certain things up. Seeing them do so in front of a classroom was really instructive and resulted in a handy list of websites and tutorials.

Finally, a great pedagogical method was the use of sticky notes (one with a sad face and one with a smiley face) which we could put up if we got stuck or had completed an exercise. This was a great example of making use of the knowledge of the crowd: participants who were successful could help out those who had some trouble with a particular exercise. And who would've thought? It turns out that having to explain something is a nifty way to see if you *\*really\** got it.

### **Intellectual and professional opportunities provided by the Institute**

This NEH Institute has been a transformative experience. First, it brought together a highly skilled group of people (instructors and participants) with a wide range of expertise and just one goal: teaching people to do better digital scholarly editing. In addition to the official program, I learned a huge deal from daily interaction with my peers who came from different backgrounds and had different levels of experience, but we recognized in each other a similar mindset and desire to learn more. We quickly formed study groups to continue working in the evening hours, focusing on topics we wanted to know more about or dealing with challenging issues we had faced during the day.

For me, the greatest opportunity provided by the Institute builds upon the concept of "students become teachers": apart from the fact that the Institute's material is all available online and can easily be reused in other workshops, I have been given the opportunity to become an instructor in the follow-up NEH Institute (summer 2019; grant application pending). The topics, but especially the computational way of thinking have turned out to be beneficial for my daily work at the R&D department of a research institute, with regard to how I structure my research as well as my discussions with software developers.

### **Scheduling and logistics (academic and practical)**

The practical logistics of the Institute were all wonderfully taken care of: from the lodgings to the daily lunches and coffee (with cookies!). As with any big undertaking (and an Institute of a full three weeks is *\*big\**) there were some minor glitches but the Institute's assistants were extremely helpful and resourceful. I enjoyed the idea of collaborative note-taking, although the technology wasn't capable of handling so many users at the same time and often got stuck in freeze mode. Nevertheless, there was email and other online chat services in addition to real life talk, so we had plenty of ways to exchange information.

As mentioned above, the schedule of the Institute was filled to the brim. It was unavoidable that certain topics got less attention than others and overall I feel the instructors made the right selection. Still I would emphasise the idea of a computational pipeline at the start of (and throughout) the Institute because it helps the participants in keeping an overview of where the Institute's going and may prevent them from feeling overwhelmed. To this end I would also strongly recommend starting each day with a road map and concluding each day with a (brief!) recap and a roadmap for the following day. This may seem redundant but I believe it's still very effective considering the amount of information that participants have to process each day. I also enjoyed learning to work with eXist DB of which I had heard a lot but never had the chance to use myself. In future endeavours I would devote more time to explaining this framework to participants, as it was not always clear what eXistDB really was and how it fitted the computational pipeline. Finally, I could see a place for the topic of project management in the Institute's program. Creating a digital edition is in fact managing a (small) project, and although the instructors did place emphasis on the planning phase, project management is a profession in itself and researchers could benefit from a professional take on the subject.

## Francesca Giovanetti

The NEH Institute has been an invaluable opportunity for me as a postgraduate student. Since then, I have taken part in my first national conference as a speaker. I presented a paper about the digital edition of Paolo Bufalini's notebook, now practically ready for publication, despite some issues with the server hosting it (this is the reason why I haven't shared the link with you yet, I will do so asap). The edition uses RDF, whose adoption I had the opportunity to discuss with the teachers at the Institute. There is a youtube video of my conference speech (<https://www.dropbox.com/s/6zg7mwax8dbm7ij/aiucd2018.pptx?dl=0>) and I am drafting an article about (I will let you know if and when it gets published). Soon after the Institute I joined EADH as their communication fellow (I am still covering the role). Moreover, I just obtained a 1-year grant to develop an open access, user-friendly tool for the conversion of digital scholarly editions from TEI to RDF expanding the work that has already been done on this topic.

I genuinely appreciated the collaborative atmosphere of the Institute as a learner.

The bootcamp was funny and absolutely necessary, but perhaps a bit too long given the total amount of time available for the summer school. I felt we didn't have enough time in the third week to discuss publication methods and issues, although you gave us a great amount of high quality training considering the time available!

The Institute definitely influenced the way I work with digital editions. The idea of thinking about the production of an edition as a "pipeline" has considerably improved my approach to DSE development. Being open since the very beginning of production is another concept I plan to apply in my day-to-day work, although this can be difficult when working with a team. For example, I wasn't able to apply this principle in the Bufalini's notebook edition. For years the edition hasn't been published because I wanted to do too much with it. We

discussed this sort of problems thoroughly during the course and this encouraged me finalizing the work by giving the right limits to the scope of the edition. I will definitely carry this lesson with me in every future project, including the one I am about to undergo.

## Paul Hackett

*What was your general experience as a learner during the Institute? If you participated in the bootcamp, we would be particularly interested in any specific comments you might have about it.* I thought the Institute as a whole was excellent, covering a wide-range (sufficient but not excessive) of techniques and tools in cutting-edge Digital Humanities.

I did not attend the bootcamp.

*How have the digital edition concepts and skills introduced in the Institute influenced the way you conceptualize, design, and implement your own editions?*

The concepts and skills introduced forced me to re-evaluate the ad hoc and haphazard way in which I had been implementing certain DH technologies, and made me re-think issues such as streamlining processes, and give thought to issues that I had not such as versioning of data and documentation of software versions in data files produced (to enable replicability, etc.).

In addition, the Institute inspired me to re-think my approaches to solving certain problem that I was facing by using new techniques and technologies introduced in the Institute.

*As we wrote last fall, we are maintaining a page of links to digital edition projects (whether completed or in progress) by members of the Institute at [https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant\\_editions.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant_editions.html), and we would like to include information from as many participants as possible. We would be grateful for information about any edition projects you are pursuing, and about how your work at the Institute has contributed to those projects. And if your edition (whether complete or in progress) has a web presence, please let me know and I'll be happy to link to it (and to update any references that are already there).*

I maintain a searchable database of primary texts and bibliographic records for the Tibetan Buddhist canon, and have been re-designing it in light of the new technologies that I was exposed to at the Institute. I hope to deploy the new version by the summer of 2019. The current version is available at:

<http://databases.aibs.columbia.edu/>

*How have you used the skills and materials from the Institute in your own teaching, either in the classroom or in collaborative projects? If you have developed a course description or project guide that was influenced by the Institute, we would be grateful if you would share it with us.*

The Institute inspired me to begin documenting the techniques that I use for problem-solving and conceptualizing projects in Digital Humanities. It is my intention to eventually teach this as a course, and possibly publish a "Best Practices in Asian Digital Humanities" textbook to accompany the course.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

The Institute has reaffirmed for me the decision to pursue academic employment specifically aimed at the digital humanities field.

*We would be grateful for any feedback you might be able to share about the scheduling and logistics (both academic and practical) of the Institute.*

Although three weeks can be difficult for some people to allocate, I think the mid-summer schedule worked best.

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

Although I was easily able to come up to speed on the requirements of the Institute for the final two weeks (I did not attend the bootcamp), it might be advisable to make the bootcamp mandatory (or at least, "strongly recommended") in order to serve as a refresher course for those already familiar with the basic concepts, but also to "debug" and idiosyncratic issues with attendees' individual computers and software installations prior to the start of the main Institute.

## Brian Long

I have participated in a number of summer programs, both under the aegis of the NEH and other programs, and the "Make Your Edition" Institute was easily one of the best I have attended, and has had a substantial impact on my research and my career choices.

As a learner in the Institute, I was consistently impressed with the quality and comprehensiveness of the instruction. Where many programs like the Institute either tend to pitch things at an introductory level or exclude beginners, I thought the Institute did an excellent job of covering introductory material while also progressing to advanced subjects. In the few cases where instruction progressed a bit too rapidly, the accessibility and helpfulness of the instructors proved to be a significant asset (and here the bonus session on XQuery on the last Thursday night with Leif was particularly outstanding). I admired the fact that the Institute did not skimp on technical and theoretical subjects (in the latter category, I found the theoretical discussion of markup languages to be particularly strong), while also attending to the practical realities of project planning and deployment.

As I mentioned during the Institute, I had already been involved in a project to produce a (physical) multilingual edition of one of the main texts that I work on. While my philological work on this project has continued, the Institute has made me think that such an old-fashioned approach in a period of technical innovation is something of a missed opportunity. More substantially, the Institute as a whole—and discussions with Mike Kestemont in particular—have inspired me to begin developing a corpus of medieval Latin medical and scientific texts, so that students and scholars can begin applying computational tools to questions of textual attribution, literary style, and conceptual changes. This is something that I long realized was a scholarly desideratum, but the Institute helped me to think about how to approach the daunting practical challenges of this kind of undertaking. Upon reflection afterwards—and not to suggest that the Institute move from its existing strengths—projects on non-Western subjects and languages were rather absent; it would have been interesting, for example, to gain some perspective on how editors of right-to-



left scripts or CJK languages negotiate the challenge of using tools that have been almost exclusively designed with Western languages in mind.

Because I've had a research postdoc at the Pontifical Institute of Mediaeval Studies for the past year, I haven't had an opportunity to formally teach any of what I learned at the Institute, but that is not to say that I have been reticent to spread the word about the Institute. My talks at the Pontifical Institute, at the University of Rochester's Robbins Library, and at the International Congress of Medieval Studies this past year were heavily influenced by the Institute, and I even led a modest Digital Humanities study group at PIMS as well.

The bigger influence of the Institute, however, is that it's inspired me to rethink my immediate career ambitions. Though I'd been thinking of staying on the history job market indefinitely, the Institute has made me realize that I'd be frustrated by a history job with no room to keep pursuing my technical interests. But rather than confine myself to academia, I've decided to explore data science for a few years, and I just completed a fellowship in Data Science at Insight (more information at [www.insightdatascience.com](http://www.insightdatascience.com)), which was quite enjoyable, and where many of the skills I learned during the NEH Institute stood me in good stead.

Additionally, I found the instructors and participants to be both congenial and intellectually stimulating, and I've continued to have a warm relationship with several people I met at the Institute. The only suggestion I'd have for the Institute in future—and it is indeed a minor one—would be to see if there's any way to have other dining options than the Pitt food court.

So in short, keep almost everything the same—my conversations with colleagues afterwards make clear that there would be a strong demand for the Institute's next iteration.

## Fernando Nascimento

*What was your general experience as a learner during the Institute? If you participated in the bootcamp, we would be particularly interested in any specific comments you might have about it.*

I did not participate in the bootcamp, but my overall impression of the following weeks was terrific. The classes were well-organized, the content was relevant to my needs. and it covered new topics that I did not know.

*How have the digital edition concepts and skills introduced in the Institute influenced the way you conceptualize, design, and implement your own editions?*

We used several concepts discussed about xml and text analysis to enhance our Digital Ricoeur project ([digitalricoeur.org](http://digitalricoeur.org)). For instance, we created a DTD specific for the Digital Ricoeur project, and we built a set of xml verification tools to check for consistency and assist the editor in creating the tagged version.

*As we wrote last fall, we are maintaining a page of links to digital edition projects (whether completed or in progress) by members of the Institute at [https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant\\_editions.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant_editions.html), and we would like to include information from as many participants as possible. We would be grateful for information about any edition projects you are pursuing, and about how your work at the Institute has contributed to*

*those projects. And if your edition (whether complete or in progress) has a web presence, please let me know, and I'll be happy to link to it (and to update any references that are already there).*

Yes, please, our project is located at [digitalricoeur.org](http://digitalricoeur.org)

*How have you used the skills and materials from the Institute in your teaching, either in the classroom or collaborative projects? If you have developed a course description or project guide that was influenced by the Institute, we would be grateful if you would share it with us.*

I have created and taught a course on Digital Text Analysis at Bowdoin College last semester. Please, find the course description attached.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

Indeed, it helped me with my teaching postdoc activity here at Bowdoin as I mentioned before.

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

Regarding improvements, my main suggestion (and it is a selfish one) would be to have a little bit more of text analysis sessions. The one we had were extremely helpful. But overall, I appreciated the distribution of contents.

## Andrea Nichols

*What was your general experience as a learner during the Institute? If you participated in the bootcamp, we would be particularly interested in any specific comments you might have about it.*

- My general experience was good, but it was a bit overwhelming. So much good information and training in such a short time. But that is what happens when it's a 3-week course.
- I participated in the bootcamp. The regular expressions, command line, Python, and error messages portions were useful. It was a little frustrating that the Windows operating system instructions for how to install and use some of the programs (e.g. Anaconda) later led to issues that meant we had to delete and reinstall. This trend of issues with software (especially for Windows users, with no alternate software options) continued during the Institute, and while frustrating, was also a very useful pedagogical moment. It gave several lessons I am applying to my DH course syllabi:
  - First, no matter how skilled and familiar you are with downloading and operating software, always remember who knows what it will be like for another person given their device and software. Have alternate software or installation options, given issues will arise. Also, be sure to leave time for troubleshooting after installation, and separate installation time from lectures on how to use it, so that students don't have to solve problems while simultaneously trying to pay attention to a lecture.

- As a result, I am scheduling troubleshooting time slots into all of my DH course syllabi, along with separate class meetings for installation, training, and then software application to their own projects to ensure the knowledge sticks. If I get to teach the course, I'm sure issues will arise from my organization decisions, but given the troubleshooting hiccups I saw within both the UNL Digital Archives and Editions course I took and in the 2017 Institute, I'm simply trying to learn from those for my own courses. However, the topics and skills from the Institute were needful, and hadn't been included in the UNL course, so I am trying to strike a balance.

*How have the digital edition concepts and skills introduced in the Institute influenced the way you conceptualize, design, and implement your own editions?*

- I continually refer back to Leif's advice to start with your most complicated sample text because all the issues of digitizing/coding will have been worked out before more texts are added.
- Also, in having seen all the software and steps necessary to create an edition **on my own**, I know how to lay out the steps for my project and feel as though I can do it myself. This is very needful, given conference presentations from major projects always imply (or explicitly state) that a \$5 million grant and a team of 10 people are an absolute necessity for any DH project, archive, or edition.
- For instance, as the scholar and 'expert' for my editorial project, I already knew the things I wanted to have the edition be able to do analytically. However, having to see the software layers necessary to make that happen helped me organize what to do with the texts once they were transcribed into XML (and what could be done even if they were raw text files), as there was a large gap for me (prior to the Institute) between transcription and the text being "on the screen" ready to use in its final form. I am also aware of so much more software to do the work, and ways to visualize the edition.
  - I know the **interface** needs to allow users to not only access and search the texts, compare multiple texts (more than 2) with more than a split-screen view, but also integrate the reader marginalia into the textual transcription with images.
  - The **API** will need to have input from users (suggesting bibliography sources, correct or add transcriptions of texts or marginalia) and let users request more than one text in order to compare.
  - Each archival book, edition of a title, author, historical figure, owner/user, and image will need a unique **UPI** in order to allow a variety of connections to other libraries, archives, projects, and databases, in addition to analysis that the edition needs to do.

*As we wrote last fall, we are maintaining a page of links to digital edition projects (whether completed or in progress) by members of the Institute at [https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant\\_editions.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant_editions.html), and we would like to include*

*information from as many participants as possible. We would be grateful for information about any edition projects you are pursuing, and about how your work at the Institute has contributed to those projects. And if your edition (whether complete or in progress) has a web presence, please let me know and I'll be happy to link to it (and to update any references that are already there).*

- My project does not yet have a digital presence, but once it does, I will happily send that along.
- I am pursuing a genre study of Renaissance-era printed English histories and their early modern users, which comes out of my dissertation work. It builds upon texts already made open-access in the EEBO-TCP partnership, along with others not yet transcribed (<http://www.textcreationpartnership.org/tcp-eebo/>).
- Between an editorial course at UNL and the Institute, I have the skills in XML and Python to get the texts transcribed and begin pulling out analytical results. The Institute in particular has been crucial for Python, XQuery, and R stylo knowledge, along with readings, visualizations, and pointers directly related to humanities work and a digital edition. Prior to the Institute, I had no clue where to go after the XML stage (other than XSLT). Knowing which software or coding languages to turn to for each stage of the project has been crucial knowledge I gained from the Institute.

*How have you used the skills and materials from the Institute in your own teaching, either in the classroom or in collaborative projects? If you have developed a course description or project guide that was influenced by the Institute, we would be grateful if you would share it with us.*

- As mentioned in my response to the first question, I am finishing up a syllabus for a Macroanalysis course (based largely on what I learned from the Institute), and I want to write a few outlines or guides for running an afternoon workshop or Hack-a-thon. There seems to be a growing need or desire to do these quick training sessions. For instance, the RSA is interested in doing transcribe-a-thons and hack-a-thons at future conferences. As someone with a DH specialty, I want to be able to join in and help, just like I did for UNL special collections after getting trained in manuscripts and rare books. Plus local universities without DH faculty and female code clubs may want to have a workshop.
- I can send the syllabus (and any other materials) along when I finish it. I am currently in the process of writing more syllabi to add to my portfolio before the fall job market begins.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

- Attending the Institute helped me when participating in an RSA conference panel this spring on the importance of the NEH (and why it needs to keep being funded).
- I was also approached to work on a digitization project for Catherine d'Medici's letters, but I think that British scholar has turned in a different direction after getting a full-time academic job. She is involved in podcasts and TV shows more now.

- The Institute has also given me the confidence and skills to want to apply for DH jobs and postdocs related to my interests, as I now feel I have enough training and knowledge. And unlike my British colleague, I am still focused on doing a digital edition out of my dissertation, no matter what kind of job I get. All I have to do is get my PhD supervisor moving along faster.

*We would be grateful for any feedback you might be able to share about the scheduling and logistics (both academic and practical) of the Institute.*

- If there was a way to move some of the bootcamp sessions to (live?) webinar videos or to training programs provided online by other DH or tech sites like *The Programming Historian*, to be completed prior to the Institute, then after each new coding language or program layer (e.g. API) was taught during the Institute, a time slot would be available for the participants to work on applying those skills to their own projects. That way, further use of the skills would happen, making the knowledge stick even more, and have the participants leave feeling a little more confident and less fearful about remembering it all.
  - I know a Python course was provided for completing prior to bootcamp, but it had so many glitches that I eventually turned to just searching for correct answers after reading each section's concept explanation, rather than learning how to do it properly myself (and I wasn't the only one who did that). Perhaps finding more reliable resources or free training courses for RegEx, Python, and command line for pre-Institute use is an area for improvement.
- For example, in the 2013 NEH seminar I attended, we not only spent a lot of time being trained in manuscripts/early print/paleography/databases, and seeing the sources available in NYC for Renaissance-era study, but we had time slots built in to revisit the archives on our own in order to apply what we learned, and report back at the end of the seminar on what we were discovering. While it was still a busy and exhausting seminar (as expected, given an NEH summer program is only 3 weeks), we felt confident going forward and had artifacts and scholarly projects to continue building upon *immediately* after leaving. After meeting up with everyone from the 2013 seminar this spring, it was amazing to see what we had accomplished in the past 5 years.
- That said, I understand why things were taught as they were at the 2017 Institute, and why it was organized as it was. It worked. The instructors were great, and all the assistants helped it run smoothly. Each person will always have their own ideas and preferences on how to organize things.
- As with the handouts, archival connections, and notes from the 2013 NEH seminar—I think an area of equal or greater value from the 2017 Institute is the website being there to refer to (which I have done, *many* times), along with all of the great materials on the "Misc" page. Also it's the connections to the people I met during the Institute, and the extensive class notes that we took, which are very valuable for future growth and help.

- Logistically, the breaks and locations for class were great. The dorms (with laundry rooms) being *right* there on campus was perfect, along with the dining hall and library being in the same vicinity. The only persistent logistical issue was the dining cards having glitches but that was an issue with the Dining Services database, and the Pitt staff was gracious in working around it.

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

See my response to the prior question for a suggestion on how to improve. Another way to think about my suggestion is to consider separating out installation from lecture, with troubleshooting time built in so when a glitch happens, installation/troubleshooting/lecture are not all occurring simultaneously. Perhaps have installation of software be an hour session after supper (when needed), in preparation for whatever is being introduced the next day. I know that means longer days (and additional exhaustion for the teachers and students), but we students were happy to stay later to learn when those sessions were offered.

## Christopher Ohge

In addition to what I wrote you last fall about my previous editions, I'd like to add some more comments to address your points in your email of 21 June 2018:

- the bootcamp at the NEH Institute (hereafter NEHI) was very helpful for me. As I've said many times, a lot of my digital training has been ad hoc, so participating in activities like the bootcamp did a lot to fill in my knowledge gaps (regex in particular).
- in teaching, I am including several lessons from NEHI: in my July 2018 London Rare Books School course on digital editing (<https://cmohge1.github.io/lrbs-digital-scholarly-editing/>), I am including sessions on digital text modelling, collateX, & lmn1, for example. In workshops for the UL School of Advanced Study's digital humanities initiative (<https://www.sas.ac.uk/projects-and-initiatives/digital-humanities>), I have also recently included command line & regular expression bootcamps based on the NEHI.
- recommendations: Some of the sessions used Python to perform XML-processing tasks that to me would have been better suited with XSLT. And since I have a working familiarity with XSLT, but still need some additional tips on more advanced processing, I think it would have been better to include some XSLT modules for processing.
- professional opportunities: you may know this, but I interviewed for my current faculty position at the University of London just a few weeks before the NEHI, and I did mention my participation in it as evidence of my enthusiasm for self-improvement :) (I imagine it did help, in any case.)

I'll add that the NEH Institute has most directly affected my most recent digital project, an edition of Mary Anne Rawson's 1834 anti-slavery anthology *The Bow in the Cloud*. This edition is not only creating a versioning text based on the original manuscripts of the anthology, but it also will include a network analysis and visualisations of variants, most frequent terms, & sentiment words. I'm also using eXist to pipe my files into a graph database (for the network analysis). All of those components were clearly facilitated by the additional training at the NEHI. This project



has already secured a small digital humanities start-up grant from the John Rylands Institute at Manchester, and a larger AHRC research funding bid is underway.

As promised, I have also attached the introduction to the special issue of *Leviathan* that I had mentioned, as well as my co-authored piece in that issue focusing on Melville's Shakespeare marginalia. The acknowledgement to the NEH Institute is on page 15 of the intro, just after "Notes".

NB. Forgot to add in the email that the R sessions at NEHI helped my marginalia text analysis publication (I mention it in the acknowledgements in the article).

## Elizabeth Williamson

*What was your general experience as a learner during the Institute? If you participated in the bootcamp, we would be particularly interested in any specific comments you might have about it.* I found the Institute a hugely valuable and productive experience, and it materially contributed to my learning and subsequent work/job role. For me, though I do have some previous experience in the practice and theory of what was being discussed, the bootcamp was essential to the workshop. It enabled us to get a uniform level of understanding across the group, and get us used to the ways of working and vocabularies that we could then build on. Completing the python course in advance was also very useful.

*How have the digital edition concepts and skills introduced in the Institute influenced the way you conceptualize, design, and implement your own editions?*

It helped me to get a wider understanding of the multiple elements one could introduce into an edition, though at times the extra narrative step between what we were learning and how it could relate to an edition was not made as explicit as it could've been. Overall, the approach of showing us a little taster of many things was very valuable, though hard work! It gave me the confidence to try new things, and I wouldn't have been as comfortable teaching myself some xquery and using eXist, as I now do in my current role, without the Institute.

*As we wrote last fall, we are maintaining a page of links to digital edition projects (whether completed or in progress) by members of the Institute at [https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant\\_editions.html](https://pittsburgh-neh-institute.github.io/Institute-Materials-2017/general/participant_editions.html), and we would like to include information from as many participants as possible. We would be grateful for information about any edition projects you are pursuing, and about how your work at the Institute has contributed to those projects. And if your edition (whether complete or in progress) has a web presence, please let me know and I'll be happy to link to it (and to update any references that are already there).*

I am currently working on an edition of the letters of Thomas Hardy, with Dorset County Museum and Professor Angelique Richardson. The edition will be launched in summer 2019. We're using eXist as our platform, and exposure to this and to associated technologies during the workshop have proved extremely useful as I continue to improve these skills.

*What kinds of intellectual or professional opportunities has participation in the Institute provided for you?*

The Institute was most useful in exposing me to a wide variety of approaches and to computational ways of thinking. I hadn't done much in the way of coding or command line work before, so it was liberating to spend a good block of time exploring and experimenting, and this in turn has been very valuable in the new position I began shortly after the Institute. Thank you!

*We would be grateful for any feedback you might be able to share about the scheduling and logistics (both academic and practical) of the Institute.*

Separating out installation time from learning time (perhaps in last half hour of previous day) would help those pesky installation delays, which really eat into the time and more importantly the flow of the sessions/days. I liked the post-it notes as a way to indicate 'all good' or 'problem here'. I was lucky to have been able to spare 3 weeks at that point in my career, but I doubt I'd get that much time away again. I'd suggest keeping it to two weeks and trimming some content. Maybe one week boot camp and one week more digital publication technologies and one week more linguistics focus, and then participants can choose 2 or 3 of them??

*Please let us know how you think we might improve a future "Make your edition" Institute—and what we shouldn't change.*

Sometimes we needed more context and descriptive introduction to explain why we were learning a particular thing and how we might apply it in the future. I think this was an effect of covering so much varied material in a short amount of time, and because the instructors were so 'at home' in it – it is difficult to take the extra step back for learners in order to explain the basics of what and why (this only needs to be brief). I think the approach to throw a large amount at us was a little overwhelming at times, and without more contextual discussion did make some in the group feel quite anxious that they weren't 'getting it', however, once we got to the end of the weeks I think most were impressed with themselves about how it had gone. All in all, I found the 'taster of many things' approach really valuable. Saying that, I'd also really value an advanced workshop specifically on eXist, Xquery and XSLT for those who already use them but could learn more – if you ever plan to run that I'd love to join!

Thanks again to yourself and all the instructors – it was a great few weeks and very inspirational!

## Mary Erica Zimmer

The Summer 2018 *Make Your Edition* NEH Summer Institute in Advanced Topics for the Digital Humanities has already exercised a substantial impact upon my work, especially in terms of possibilities now more readily within my grasp, as well as future research trajectories more concretely planned.

Chief among these is my excitement at the very real possibilities presented by corpus-based editions, which were both implied and developed by the Institute over its course. This is the area in which I wish to concentrate my digital humanities work, and the range of tools, platform possibilities, and theoretical considerations introduced by the Institute has both oriented me to means by which this vision may be achieved and reinforced my sense of a compelling research community desiring to pursue similar topics.



Three projects benefit most immediately from topics explored at the Institute: my own work detecting textual reuse in the work of Geoffrey Hill, my dissertation's subject; my work with the expanding corpus of *Digital Mitford: The Mary Russell Mitford Archive*, and projects in the early modern arena on which I have worked with the Folger Shakespeare Library and with groups at MIT, where I am currently a Research Associate. The ability to articulate in concrete, realizable terms how I plan to develop projects in all three areas has been of immense and ongoing benefit. I do wish I had been able to attend the Python Boot Camp week: while this was not permitted by my schedule, it would have been extremely beneficial, even though I did possess familiarity with the language. The muscle memory involved in work with the command line, as well as with Git and GitHub, makes orientation valuable and builds a common vocabulary among participants, both practically and theoretically. (In digital humanities, the two often overlap in surprising ways.)

I do wonder whether participants' experience of the institute might be made slightly more cohesive through reference to a shared, pre-existing project throughout. Even if such a project were simply to serve as a recurring example, this approach might prove useful, as well as enrich participants' sense of particular strategies' immediate possibilities.

My work as a Research Associate at MIT has been deeply influenced by the Institute's vision: at present, I am developing corpus-based modules within *Global Shakespeares: The Merchant Module* that are designed both to enrich students' research experiences via introductory programming environments and enhance faculty's appreciation of algorithmic approaches in humanities research. As these modules are released, I will provide links to the Institute organizers to be shared with the NEH.

As well, I will share further work with *Digital Mitford*: this project's open source commitments make it particularly ripe for illustration, instruction, and experimentation.

As I prepare to enter the job market, I am developing digital humanities syllabi oriented to editing and will most certainly draw upon the Institute's resources in this regard.

I am extremely grateful for the experience and look forward immensely to remaining in touch with the group and its projects in years to come.

## Appendix V: September 2017 participant evaluations of 2017 IATDH

Although we planned to ask for substantial feedback only a year after the conclusion of the Institute (those responses are in Appendix IV, above), in September 2017 we wrote to all participants to request preliminary baseline information about their experiences in the IATDH, as follows:

I am writing at the moment to bring you up to date about plans for the evaluation of our NEH "Make your edition" Institute. Because the translation of new methods into results often does not happen quickly, we proposed initially to the NEH that we would write to the participants approximately one year after the conclusion of the Institute to ask about

the impact it may have had on your research and teaching. In the interest of identifying a baseline for this later survey, though, I am writing now to invite you all to send a brief message (from a couple of sentences to a couple of paragraphs) about how you see yourselves using what you learned at the Institute in your research and teaching. I would be grateful if sometime in the next week (by Thursday, September 28) you could please send any information you would like to contribute to [djbpitt+neh@pitt.edu](mailto:djbpitt+neh@pitt.edu).

The full text of all responses that we received (from 13 participants) is below. These early responses show a commitment to original and innovative research-driven editions and an awareness of how to go about designing and building them. All confirm the success of the Institute as having transformed the participants' ability to model, conceptualize, and begin to implement their editions.

The last email in this Appendix is a query received from the professor of one of our participants, asking for information about how he might obtain training similar to what we provided for his graduate student. ([Email from Victor Millet, 2018-01-14](#))

In the semester following the conclusion of the Institute, one of our participants, Dr. Christopher Ohge, assumed a new faculty position at the University of London. In an interview on his University's web site (<https://englishstudies.blogs.sas.ac.uk/2017/11/03/an-interview-with-dr-christopher-ohge/>), Dr. Ohge comments on how our Institute has transformed his understanding of the theory and practice of digital editions: "echoing what I learned from David Birnbaum, I would say that digital editing is the computational pipeline (from encoding and processing digital documents to interface design) that brings a scholarly edition to a digital medium."

## Email from Marie-Claire Beaulieu, 2017-09-20

Saturday, February 24, 2018 at 10:54:57 AM Eastern Standard Time

---

**Subject:** RE: NEH Institute: Make your edition

**Date:** Wednesday, September 20, 2017 at 10:34:29 AM Eastern Daylight Time

**From:** Beaulieu, Marie-Claire

**To:** Birnbaum, David J

Dear David,

thanks so much for your message! I am so happy I had the opportunity to attend the Institute! I've already have a chance to practice some things I learned, so here are a few bullet points of things I've done since coming back. I also think that you are right that results will take a while to show their full effect, so I am excited to see what I'll be doing when we check in next year!

Since coming back, here are some ways I've used my new skills:

-For my current research project, I have a large amount of raw OCR files. I was able to rename them all quickly by the corresponding page number in the original book using the command line.

-I am co-teaching with a Computer Scientist, and for a module about data cleanup, I have made a KNIME workflow that makes heavy use of Regexps to sort the data of a particular column into categories and then normalize it. The data we are using is the "floruit" column of the Lexicon of Greek Personal Names. I plan to give my students an introduction to Regexps during that lesson (and BTW I'd like to reuse and adapt the excellent Regexps page you guys put together if that's OK)

-My colleague from CS is putting small nodes of Java in our KNIME workflows for various operations. Using what I know now of Python, I was able to follow what that code does and repurpose it in other situations.

I hope that's helpful!

Best,  
Marie-Claire

Marie-Claire Beaulieu  
Associate Professor  
Director of Graduate Studies  
Department of Classics  
Tufts University

## Email from Gustavo Fernández Riva, 2017-09-27

NEH Institute “Make *your* edition”: Course Review – September 2017

Gustavo Fernández Riva

The areas of the Institute that have the greatest impact on my own activities are computer assisted collation, stylometry, and the transformation of XML-encoded texts using Python and Xquery. In these three fields I had little to no knowledge prior to the Institute and now I understand the basics upon which to continue learning.

I am already using CollateX to assist me when collating the texts I am editing. Before the institute I had not been able to fully understand and implement this tool just by reading the documentation. The participation in the course also showed me how powerful and useful it could be. There are still some possibilities I haven't explored and I hope to be able to in the future.

I had absolutely no knowledge of stylometry before the course, and now I am thinking about many possible application for the corpus of medieval texts I research. Next year I might include stylometric analysis as part of a research project for which I will be applying together with other members of my home institution.

Finally, the lessons on Xquery and the xmltree library in Python were very useful to widen the possibilities when processing XML/TEI documents. I used to do all the transformations with XSLT, but now I have more options.

The theoretical discussion about editing was very helpful too. I now see my own edition in a different light. Seeing projects from other colleagues and talking about them was very helpful to redefine my own goals and assumptions on digital editing.

Finally, I will probably teach a graduate course on digital humanities next year and will include some of the contents seen during the institute, specially the nltk library in Python and stylo in R.

## Email from Halila Bayramova, 2017-09-28

NEH 2017 Summer Institute Make Your Edition -- Preliminary Report

Halila Bayramova  
2<sup>nd</sup> year PhD student  
Trinity College Dublin

"Towards a Digital Genetic Edition of *Finnegans Wake* Chapter X: Night Studies"

In terms of its impact and utilization potential the NEH 2017 summer institute "Make Your Edition" could not have been more aptly tailored for my doctoral research project. A genetic critical approach adds another dimension to the edition-in-progress and makes it a special case even among digital scholarly editions. Thus, the project demands its own questions and requires quite customized answers.

Since the institute covered almost every(!) aspect of my research, it might be more practical to point out the instances I found particularly invaluable.

### 1. SAFETY FIRST

One of the highlights of the workshop for me was the special emphasis on preservation, sustainability, and adaptability of digital scholarly editions. Attempting to surpass the boundaries and reliability of the print by employing such a dynamic and volatile medium as the digital is always daunting for scholars. Though the new medium's potential is boundless, an editor cannot (& should not) ignore the pertinent and constant concerns over resilience, longevity, and endurance of editions. What the institute had to offer us was the state-of-the-art practices for a range of different on-going and planned digital projects.

### 2. GITHUB

GitHub as a workstation, updatable repository, and a dissemination point had never been on my research radar for a couple of reasons, the chief one being the severe copyrights restrictions of Joyce manuscripts. The NEH institute was very helpful to highlight the advantages of GitHub and how it might enhance my research experience from the very early stages without infringing Joyce's copyrights.

### 3. LEGAL RUMINATIONS

I found the seminar on copyrights in the digital realm very practical since the legal significance and consequences of my project necessarily require considerable dabbling in the convoluted and often contradictory world of Joyce copyrights. But what I had failed to fully register and what was a revelation to me at the Institution was the two-way nature of the legal part of many digital projects. A full consideration should be given not only to the permission for utilization of the *inbound* textual material but also to the protection of the *outbound* data comprising of -- in my case -- the very data plus the editor's code.

### 4. USERS

Thinking about the audience (readers) of "Night Studies" DGE during the Institute sessions caused a slight shift in this aspect of my research. If the dissemination frame used to be narrow and the mindfulness of audience was a very late-stage consideration, I've been gradually learning to be conscious of the final product even as I transcribe the manuscripts, since it affects many conscious choices the editor has to make from very early on.

### 5. SOCIAL ASPECT

On a more humane note, besides the technical expertise and a vast range of skills, the summer institute provided an incredibly unique opportunity to meet kindred minds working in similar or relative fields, facing the same sort of problems, asking similar questions, and voicing familiar concerns. It was very uplifting to experience that your work mattered and that people shared the same enthusiasm and interest in the digital humanities. Needless to mention, such an experience also created a wide network of colleagues and friends.

N.B. I especially appreciated the instructors' advice on how to search and find answers to our questions rather than providing us with answers themselves.

## Email from Paul Hackett, 2017-09-20

Saturday, February 24, 2018 at 10:53:23 AM Eastern Standard Time

---

**Subject:** Re: NEH Institute: Make your edition

**Date:** Wednesday, September 20, 2017 at 2:09:24 PM Eastern Daylight Time

**From:** Paul Hackett

**To:** Slavic - NEH

Dear David,

Thank you once again for what was a thoroughly enjoyable workshop and educational experience.

In terms of my current anticipated use of what I learned at the Institute, I would say:

- (1) A clear understanding of the necessity of embedding versioning documentation not only in software, but in the data produced by them for tracking and replication of results, and, by extension, a solid protocol for correlating software and data documentation. I have already begun implementing this in my new projects and retroactively applying it to legacy projects.
- (2) A well-grounded understanding of different tagging schemes (XML, TEI, etc.) and various software packages for manipulating such data. I have already begun modifying my data processing routines in light of what I learned, and envision eventually reconfiguring large portions of my public databases to utilize XML data.
- (3) Different types of visualization of data sets and different aspects within them, as well as a basic understanding of the theory behind the visualizations. I envision adding such visualization functionality to both the research and public interfaces to my projects.
- (4) A broad overview of the history, theories, and practices in the field of digital humanities. I have begun systematically re-working my notes from the institute into a general guide and hope one day to teach a university-level methodology course on this subject.

----

All the best,

Paul

## Email from Ming Yeung Cheung, 2017-09-28

Saturday, February 24, 2018 at 10:33:05 AM Eastern Standard Time

---

**Subject:** Re: NEH Institute: Make your edition  
**Date:** Thursday, September 28, 2017 at 1:49:13 PM Eastern Daylight Time  
**From:** Ming Yeung B. Cheung  
**To:** Birnbaum, David J

Dear David,

I'm doing well. I hope this email finds you well, too!

I would like to take this opportunity to thank you again for organizing this workshop, and for the partial reimbursement of my travel expense, for which I received a check some time ago.

In terms of research, I have submitted a proposal for funding to enhance an ongoing project in our department, an "Online Database of Religious Sites", by designing a TEI tagging schema to markup the textual descriptive data of the entries. This will improve the usefulness of the digital archive by providing richer information beyond mere statistics. Moreover, I intend also to work with a software engineer to design APIs for the database so that other researchers can assess and use the data that our team of surveyors have collected for further research. In the meantime, I will continue to reflect on what kind of edition should we make for our Digital Ricoeur project.

I am also planning a talk / workshop for my colleagues in the college of social science on what we have learned in the NEH Institute in Pittsburgh. I will put the emphasis on the project-planning part (goal-oriented, pipeline model, version control), coding (Python, Jupyter notebook), publication (database-server-client model, APIs, RDFs, Linked Open Data) and some basic ideas of text tagging (TEI, XML, etc.). As for text analysis, some topics like collation may not be that interesting for social scientists (economists, psychologist, and sociologists) but researchers in history and literature may be more interested in them. I will make a proposal to the college dean so that she can decide what should be included.

I hope this is what you expect, and if you need any clarification of what I have written here, please let me know.

I look forward to report my progress in a year's time!

Best,  
Ming Yeung



## Email from Nicola Reggiani, 2017-09-29

---

Saturday, February 24, 2018 at 10:30:10 AM Eastern Standard Time

**Subject:** Re: NEH Institute: Make your edition  
**Date:** Friday, September 29, 2017 at 8:25:33 AM Eastern Daylight Time  
**From:** nicola.reggiani@gmail.com on behalf of Nicola Reggiani  
**To:** Birnbaum, David J

Dear David,

nice to hear from you! Hope everything is well! Sorry for the slight delay in my answer... anyway, I would just say that certainly I'm planning to apply many things I learned at Pittsburgh to improve my current project of the digital database of the Greek medical papyri, in particular as regards the GIT repositories to manage our data (a GIT already exists for the papyri projects, but it would be nice to develop our own data repositories at some point), the improvement of regular expressions and the application of XQuery / XPath technologies to allow for deep searches in the corpus. I still have to figure out many things but I think that these will be the main outlines.

All best wishes and thanks again for everything!

Nicola

## Email from Les Harrison, 2017-09-21

---

Saturday, February 24, 2018 at 10:51:10 AM Eastern Standard Time

**Subject:** Re: NEH Institute: Make your edition  
**Date:** Thursday, September 21, 2017 at 9:26:07 AM Eastern Daylight Time  
**From:** Les Harrison  
**To:** Birnbaum, David J

Hi David -

The NEH "Make Your Edition" summer institute has led me to rethink my Poe Edition. I am putting more emphasis on MS Acquisition and have created a small, three-person team to research previous Poe editions as we create a more developed conception of this edition.

As a result of the institute I am now working with a faculty member in my department to create a new course, "Editing the Critical Edition" which will lead students through the creation of small digital editions of scholarly texts. If this collaboration is successful, we hope to make this course a regular offering in our graduate program.

Hope the fall semester is off to a good start. All the best -

Les



## Email from Rikk Mulligan, 2017-09-29

Saturday, February 24, 2018 at 10:26:01 AM Eastern Standard Time

---

**Subject:** Re: NEH Institute: Make your edition

**Date:** Friday, September 29, 2017 at 11:25:45 AM Eastern Daylight Time

**From:** Rikk Mulligan

**To:** Birnbaum, David J

Hi David,

My apologies for missing the initial deadline.

The NEH "Make your edition" Institute offered a broad array of methods, tools, and approaches to creating, analyzing, collaborating, and sharing a digital edition online. As a researcher, the briefings and instruction on command-line tools and text-analysis have already informed plans for creating a corpus and scoping out needs for an environment, platform, and tools needed for analysis. The specific focus on sharing both the datasets and the analysis will help plan for interacting with different groups of researchers, as well as the public.

As core faculty in a nascent digital scholarship center, the organization of the institute as a whole offers me and my time points of comparison for how we collaborate with our local scholarly community and plan for projects and workshops. The way the institute structured its sessions, including organizing content from a basic tool bootcamp through specific, deep methods foci suggests how we might offer short workshop sessions, day-long workshops, and mini courses that work along similar lines. The institute as a whole provided a model for collaboration that will help us remind our students and researchers that they are part of a global community and not limited to the resources and skillsets available on our campus.

Rikk Mulligan  
Digital Scholarship Strategist  
University Libraries Faculty  
Carnegie Mellon University  
email: rikk@cmu.edu  
phone: 412.268.7315

## Email from Fernando Nascimento, 2017-09-30

---

Saturday, February 24, 2018 at 10:24:03 AM Eastern Standard Time

---

**Subject:** Re: NEH Institute: Make your edition

**Date:** Saturday, September 30, 2017 at 8:53:03 AM Eastern Daylight Time

**From:** Fern Nasc

**To:** Birnbaum, David J

Hi David,

I hope you are doing very well.

Sorry for the delay in getting back to you.

As I mentioned, this semester I am teaching at Bowdoin College in a postdoc position. This fall I am teaching a course entitled "Computation in Context" in which I am adding several aspects of what we discussed during the institute, especially XML formatting, stylometry and a couple of tools such as jupyter notebooks and doxygen. Next semester I will be teaching "Digital Textual Analysis" and these components will be even more present. I am also planning to teach it in Python.

Regarding the Digital Ricoeur project, we are preparing to present it to the Society of Ricoeur Studies during its annual conference that will take place in Boston next week. We are currently using a relational database, but I am considering the possibility to migrate to a XML-driven one such as exist db.

Best Regards,  
Fernando.

## Email from Francesca Giovannetti, 2017-10-01

---

Saturday, February 24, 2018 at 10:21:01 AM Eastern Standard Time

---

**Subject:** Re: NEH Institute: Make your edition

**Date:** Sunday, October 1, 2017 at 3:02:17 PM Eastern Daylight Time

**From:** Francesca Giovannetti

**To:** Birnbaum, David J

Dear David,

Apologies for the delayed reply!

As a young graduate, participating in the Institute brought new opportunities for me. I have been elected Communications Fellow for EADH (European Association for DH) and I am preparing a proposal for the upcoming AIUCD conference (Italian Association for DH). The content of the proposal, entitled 'Linked data and digital scholarly editions: the case study of Paolo Bufalini's notebook', stems from the new skills I have acquired at the Institute. If the proposal is approved, I will have the chance to take part to my first national DH conference as a speaker. I also hope to find enough resources (time and money!) to complete my work on the edition. In any case, I will surely upload all the materials I have prepared so far to GitHub.

Best wishes,  
Francesca

## Email from Albertina Hughey, 2017-10-04

Saturday, February 24, 2018 at 10:16:35 AM Eastern Standard Time

---

**Subject:** RE: NEH Institute: Make your edition  
**Date:** Wednesday, October 4, 2017 at 2:22:33 PM Eastern Daylight Time  
**From:** Hughey, Albertina  
**To:** Birnbaum, David J

Good morning, David—

I miss you! We got off to a late semester start here in Texas because of storms, so I'm still swamped with managing student/teaching matters.

But I arrived home and immediately began re-studying our NEH Institute materials. What's more, I marched it all over to the Computer Science Department where I'm now auditing three courses—Data Files & Structures, OOP C++, and Theory of Computation (Kestemont's algorithms writ larger)—to fill in the gaps between what I learned this summer and what I did not know before arriving.

Python is sooooo much more lovely than C++ (and faster for building application parts), but I'm keeping an open mind re: using C++ to extend Python for built-in objects. If all goes to plan, I'm on track to roll out a prototype of my digital critical edition by late Spring (our Paul's agreed to coach my API building and testing).

Equally exciting for me right now is sharing the knowledge and technologies we explored this summer, emphasizing the democratizing nature of linked-open data & technologies, with students and colleagues. Workshops are on the horizon, and I'll keep you posted on everything as it all unfolds. Thank you again for everything!

Albertina

*Albertina L. W. Hughey, Ph.D.  
Associate Professor  
Department of English  
Texas Southern University  
MLK, Jr. Humanities Bldg., Suite 145  
3100 Cleburne Street  
Houston, TX 77004  
Phone (713) 313-1359  
FAX (713) 313-7598  
hughey@tsu.edu*

## Email from Christopher Ohge, 2017-10-05

---

Saturday, February 24, 2018 at 10:09:51 AM Eastern Standard Time

---

**Subject:** RE Make your edition

**Date:** Thursday, October 5, 2017 at 6:44:53 AM Eastern Daylight Time

**From:** Christopher Ohge

**To:** Slavic - NEH

Dear David,

I'm sorry this response is coming in late. I was in the middle of moving for my new job at the University of London & I've just started to settle in a bit. If my thoughts still might be of use to you, here they are:

I am planning to use what I learned at the Institute for research & teaching. In my research, I am currently building on what I learned about R & stylometry for a project on word frequencies in Herman Melville's reading & marginalia. I also plan to work more on Python & CollateX to create a digital genetic edition of Mark Twain's *Pudd'n'head Wilson* (possibly, also, with the aid of eXist DB). Tho' this was not explicitly covered in the Institute, it became clear to me that I needed to continue to educate myself in XQuery & Python to streamline my editing workflow.

As to teaching, I will incorporate what I learned during the bootcamp portion of the Institute for digital literacy training & file management. I am also teaching a textual scholarship course in the winter that will include some of the material from the Institute -- e.g., the principle of computational pipelines, using alternative markup such as LMNL, & creating visualizations from edition data.

With best wishes,  
Christopher

## Email from Brian Long, 2017-09-25

Saturday, February 24, 2018 at 10:41:59 AM Eastern Standard Time

---

**Subject:** NEH early response

**Date:** Monday, September 25, 2017 at 12:04:58 AM Eastern Daylight Time

**From:** Brian Long

**To:** Slavic - NEH

Dear David,

Thanks for the email following up on the NEH Institute. Things here in Toronto are, indeed, quite well: I've found a quite pleasant place to live, and am finding it very easy indeed to get a lot of work done with PIMS' fantastic library. I'm still wrapping a couple of articles up from the summer at the moment, but should be on to new work soon (including getting back to text editing). I hope all is well back in Pittsburgh, too!

I've written up a brief description of the direction I'm planning to take my project in below, but I wanted to say that I really am quite interested in more seriously pursuing the project I presented on at the Institute. There were several projects and lines of research that I had figured were feasible in some way (and that I had come up with kludgy solutions for), but that the Institute showed much better, more robust ways to implement (as I describe below).

I wanted to ask, though, if you had any suggestions— or even just general advice— about finding funding to undertake a DH editing project of the kind I'm envisioning. At the Institute, as it happens, I ended up spending more time talking to the European participants and instructors about the possibilities for funding in Europe than I did those from the US. There are a number of postdoctoral fellowships here in the states, of course, as well as the NEH Digital Humanities Advancement Grants, but I'm a bit unsure about how enthusiastically a project that is largely textual (though cross-linguistic) will be received, especially from a junior scholar with no long-term institutional affiliation. Perhaps I'm envisioning too much resistance and should just cast my bread upon the waters, but really any suggestions you'd have would be quite appreciated.

And again, thanks so much for the Institute; it was enormously stimulating and rewarding, both for conceptualizing my project, for getting a clearer technical sense of how to implement it and learning about tools I was unaware existed, and especially in getting to meet all of the participants and instructors. It was both informative and invigorating.

Warm regards,  
Brian

### Preliminary Project Statement

During the "Make Your Edition" NEH Institute, I developed a familiarity and facility with a number of tools and approaches that I plan to draw upon in both my research and my teaching.

As far as my research is concerned, I came to the institute with a vague idea that it might be interesting to produce a digital edition of the tradition of a text that has been central to my research up to this point, the 'Viaticum' of Constantine the African. Over the course of the Institute, however, I realized that what I was learning had opened up instead a much wider set of possibilities, and that an entirely different project— that I had vaguely entertained, but put off for technical reasons— would be feasible. In particular, I realized that by drawing upon what we had learned about the Gothenburg Model's pipelined approach to collation, it would be relatively straightforward to modify the CollateX collation tool to automatically align Arabic medical texts and their Latin translation. This kind of automated alignment, in the long run, would facilitate the production of a lexicon of Arabic-Latin medical translations, something that has long been a scholarly desideratum, but which has remained out of reach because of the considerable number of texts in question. While at the Institute, I produced an initial version of my Arabic-Latin alignment tool, and began to fine tune it on the basis of the text I know best, Constantine the African's 'Viaticum' and its Arabic source text. After the Institute concluded, moreover, I have been working on modifying it to accommodate many-to-one and many-to-many matches as well.

This project was only one of the fruits of the Institute, however, and there are a number of other projects I have realized the feasibility of; most notably, this includes the use of digital stylometry for medical texts of uncertain attribution and textual reuse in Greek medical sources. In the nearer future, however, I can well imagine that I will draw upon the material from the Institute in teaching both how to use the tools we worked with and broader methodological issues, such as, for example, realizing the limitations of the data model of your tools.

## Email from Brian Long, 2018-02-05

---

Friday, February 23, 2018 at 12:40:19 PM Eastern Standard Time

---

**Subject:** Digital Projects Update

**Date:** Monday, February 5, 2018 at 1:30:59 AM Eastern Standard Time

**From:** Brian Long

**To:** Birnbaum, David J

Dear David,

Apologies for taking so long to send you this, but I thought you might like to know that the little translation alignment tool I put together during the Institute is up on my Github repo, at <https://github.com/mrgah/tralalign>

It took a while to get to a good state because I went down an odd rabbit hole: I wrongly assumed setting CollateX's output to html would return some kind of HTML object I'd be able to output to a file, when in fact it doesn't return any kind of object at all, but just displays HTML (in an IPython notebook, for example). At any rate, I've taken it back to its IPython notebook roots so that other people can use it if desired.

I've been working away on my big editorial project here at PIMS, but I've also been working to develop a corpus of medical and scientific texts that I plan to use for digital stylometry, text mining, etc. (These are texts that have often been excluded from the larger textual databases.) It's a private github repo (mrgah/med-stem) at the moment (some of the texts in the corpus are still in copyright), but I thought I'd let you know that it's under active development. The description I've got is this:

"A github repository of a textual corpus of medieval scientific and medical texts. Meant to be inclusive, this includes both works in the fields themselves as well as works that draw heavily upon scientific and medical materials."

It's heavily influenced by the Institute, (conversations with you, Mike, and Paul Hackett in particular) so I'd be happy to provide any further information that might be helpful to you.

Thanks for everything-- the Institute has been hugely beneficial to my work and to me, and I'm extremely grateful to all of the hard work you and the other instructors put into it.

Warm regards,  
Brian

## Email from Victor Millet, 2018-01-14 (Gustavo Fernández Riva's professor)

Friday, February 23, 2018 at 12:44:05 PM Eastern Standard Time

---

**Subject:** Institute 2018

**Date:** Sunday, January 14, 2018 at 6:44:06 AM Eastern Standard Time

**From:** Victor Millet

**To:** Slavic - NEH

Dear professor Birnbaum,

I am professor for medieval literature at the University of Santiago de Compostela, Spain. A PhD-student from Buenos Aires who is under my supervision, Mr. Gustavo Fernández Riva, attended the Advanced Institute in the Digital Humanities last summer. I would very much like to follow a similar training this summer. But as far as I could see at the Pittsburgh-NEH website there is no such course planned for this year. Could you please confirm this? And could you perhaps recommend me similar hands-on trainings at other North-American or European universities?

Thank you very much for your kindness.

Best wishes

Victor Millet

## Appendix VI: Daily feedback

As described in the white paper, during the Institute we asked the participants to answer brief daily survey questions about the materials and instruction. Instructors then adapted the curriculum to this feedback on the fly, among other things by allocating additional attention to topics that had proved especially challenging and by scheduling optional review sessions.

## Week 1, Day 1 (BootCamp)

	Yes	Somewhat	No
I understand how to explore and navigate within my file system from the command line	11		
I understand how to manipulate (copy, move, view, etc.) files from the command line	9	1	1
I understand how to create and saving aliases, and how they might be useful	4	2	5
I understand how command line switches modify the behavior of commands	10	1	
I understand how to look for more information about commands	11		

Note: some of the info simply wasn't covered on day 1, but students learned it on day 2

## Week 1, Day 2 (BootCamp)

	Yes	Somewhat	No
I understand what regular expressions are, and where they might be useful	10	1	
I understand how to edit on the command line	10	1	
I understand how to use variables on the command line	10	1	
I understand how to read commands and files	11		
I understand how to read from and write to files from the command line	10	1	

### Comments

It is certainly sticking more than the first day. It's good to be reminding us of how it was shown to us already or that 'we are now using 'X' thing that was shown' to help us being able to recall from the flood of info. When we were reminded today of that a few times, it was good. As time goes by, just be sure to have patience with us

While I learned a lot out throughout the day, I have noticed that there has been a problem communicating some of the concepts clearly. I have noticed this with other DH programs: you will often see frustration from the instructor that suggests "this is so easy for me to get, why can't you get it"? This is perhaps best evidenced by examples of, say, complex functions in one's bash profile. If you show me the function, & we want to learn how to do that, & then the instructor says, "I can't, really. It's too complicated"—that is not very helpful. You get the sense sometimes that the instructor is showing off a bit, without clearly explaining how one can get to that level of proficiency. This is potentially problematic, because it furthers the programmer / humanist divide, & makes programming more intimidating. Despite this, there were points throughout the day where the communication was clearer, but it's nevertheless something to be mindful of. (I should that I have been guilty of this myself when teaching DH to my students, so I understand the difficulty). Many thanks.



## Week 1, Day 3 (BootCamp)

	Yes	Somewhat	No
I can use regex with grep on the command line.	9		2
I know what Jupyter Notebook is, and how it may be useful	11		
I can interact with Python in Jupyter Notebook	11		
I know what NLTK is and how it might be useful	7		4
I can process a single text file with Python inside Jupyter notebook	8		3

### Comments

David and Na-Rae did a great job on their stuff. It was fast-paced, and I know we have a lot to cover, but I appreciate the willingness to slow down or stop if there are issues. David's sessions always seem to make sense to me, no matter how fast. Na-Rae's overview of Python and working on commands was very useful. I can see why we were to do CodeAcademy ahead of time, but that tutorial had issues, whereas Na-Rae's made more sense. Leif is great for weaving among the crowd to see who has issues. Feeling more confident. Definitely had fun with the file of English language words with Na-Rae!

## Week 1, Day 4 (BootCamp)

	Yes	Somewhat	No	
I understand the basics of how communication over the Internet works	9		2	1
I understand the Git workflow	10		2	
I can perform some basic operations using command line Git	9		3	
I know how to clone a Git repo	11		1	
I know how to go about processing a text corpus with Python	8		4	

### Comments

The value of NLTK makes much more sense today!!! It's nice to have the review session. A well-thought out addition to the daily cycle. Helps to have issues or questions ironed out in person, rather than a webpage or email sent every evening or morning before the next session. I know we're getting used to each other and more comfortable with the class, but if talking in subgroups gets any more frequent or louder, I will get too distracted to be able to keep up with the lecture that is going on simultaneously. I appreciate the enthusiasm, and willingness to help (too many people have a tendency to just live on their own island), but jumping in ahead of the instructors or overlapping vocally with them makes the background noise I have to filter out a bit much. My brain can only take so much when I'm so paranoid about focusing on the instructors to learn what they say, given the instructors are going borderline fast sometimes anyway, and I can't afford to miss a step. I really appreciate all five instructors working with both Windows and Mac, or providing help with both, as sometimes I can run into DH groups or scholars that are 'sun-rises-and-sets-with-Mac' snobbish attitude, and that doesn't help in training others who wish to learn about Mac, or for working in multi-device environments. Having familiarity or expertise in a single platform is different from being snobbish (or from explaining the pros/cons of each platform), so the attitude of collegiality and group effort among the instructors is a good example to set. Things are sticking!! I'm excited and happy!!

Python workshop very well managed.

## Week 1, Day 5 (BootCamp)

	Yes	Somewhat	No
I have configured my GitHub account to use my SSH keys, so that I don't have to type my password every time I push.	8		
I understand the purpose of forking and branching in Git and GitHub.	6	2	
I know what Git merge conflicts are, how to reduce their likelihood, and how to deal with them when they occur.	5	3	
I understand the role of HTML, CSS, and JavaScript in delivering web content.	8		
I know what Markdown is, and why it might be useful.	8		

### Comments

all very relevant stuff that was taught in a straightforward way.  
It has been excellent to focus on all of these topics for a full

## Week 2, Day 1 (PhilCamp)

	What excited you most about what we did today? I think starting to plan our projects was exciting, and talking it out with others gave me good ideas and feedback about the value of my project and concrete ideas to try to implement it (I am at the very beginning of a new project)	What did you find most challenging of what we did today? I was very slow with trying the xpath queries. I think I could have got there, but it was difficult to follow the rhythm at which we were going. It may have been useful to have this as homework, but I am planning to return to it after the end of the Institute Attempting to follow the excessively rapid explanations of XPath and other functions was the most challenging. I saw a number of XPath beginners get lost, and it would have helped to devote more time to reviewing some solutions together as a group. Alternatively, you might ask those experienced with XPath to work together with beginners to walk through a handful of the exercises. The exercises *should* be very helpful to those weak or rusty on XPath as a way of showing what kinds of XML information you can locate and process, but when presented rapidly they may be discouraging to those who need them most.	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu			project planning and modeling
Elisa Beshero-Bondar	The discussion of modeling and collation, and of what decisions we've made to limit what we encode in our edition projects.		Thinking about modeling and visualizing collation is helpful to me just now.
Ming Yeung Cheung	to think about our models of the text, and the problem of overlapping in XML tagging	to articulate what data I do NOT capture Without question it was XPath, which was presented perhaps a bit too quickly and at the end of the day. I hope that we return to that because I know that I'll need to understand that before I can go on with XQuery, graphic visualizations, etc. etc. I got totally lost in the XPath discussion. Very unclear on the concepts and had a lot of trouble applying the concepts during code lab.	the advice on starting from the goal the project, and taking iterative small steps
Roxanne Eberle	I'm starting understand the ways in which my TEI encoding might complicate data extraction.		XPath and its complications
Les Harrison	I thought the project planning discussion was very useful. The discussion of OHCO was interesting.	Wished we could have discussed some of the material posted to the git pages for W2D1 which ere skipped - in particular during the 11-12:30 section.	Project planning. I'll use XPath but I'm not sure today's presentations have aided my understanding.
Albertina Walker Hughey	Working on the command line. Beginning to learn XPath, and what XQuery was (heard the terms but never had it fully explained). Oddly enough, XPath really interested me and I wanted to solve my initial issues with understanding it. David really helped. Thanks!! Also, hearing "Even I had to look this up before teaching it..." is good to know, when I sometimes feel like I'm floundering around	Working on the command line.  Adding another language to my rotating rack / Merry-go-round of coding languages (they might be blurring a bit). Being on cold medicine, so my head feels detached, so it was a bit more difficult to follow along but I got it (just with 5 more min of work than everyone else...or that's what it felt like).	Working on the command line.
Andrea Nichols	The different possibilities and ways of modeling a text, including the discussion and comparison with the others' ideas.	XPath (and the quick look at XQuery): it takes a few time to understand how it works in the XML structure, but then it is a powerful tool.	Being reminded of how many XML tags I still have to look up, so I need to stick with it; also XPath finally being explained to me from the ground up. Certainly the various ways of querying an XML source with XPath, which can offer some more possibilities than the current papyrological databases and tools.
Nicola Reggiani			

# Week 2, Day 2 (PhilCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu	I enjoyed the discussion "what story are you trying to tell". It got at the heart of editing projects, in which defining the question is both crucial and challenging. I think the discussion and examples presented were helpful.	Normalization. We were going very fast, and I had trouble following the code. Also, on a practical level, it was often difficult to see the code on the board.	edition terms and concepts
Elisa Beshero-Bondar	Discussion of tokenization and alternate strategies for normalization.	Installing and running Calabash and Alexandria-- instructions could use some fine-tuning.	Tokenization and collation
Ming Yeung Cheung	Different schemes to markup texts.	Using Python lxml tree and pulidom libraries to manipulate XML/TEI text.	Using Python to handle TEI tagged texts.
Roxanne Eberle	Learning Tokenization	Learning Tokenization. I'd really like us to return to reviewing and asking questions at the end of every day. I feel as though we are getting 50% or 75% to a goal and then stopping short. Since we're moving so quickly, I think that this is impeding my retention. I know what questions to ask but not how to answer them.	I thought that the reminder to continually remind yourself about the theory of your edition was a useful one.
Paul Hackett	XML parsing	the multiple working environments & requisite software installs (and debugging)	strategies for POS tagging& lemmatization, and the XML structures for representing the output of those processes
Les Harrison	I thought the "theory of the edition" session was very useful. I would have liked a more formal discussion of types of edition instead of looking at various data visualizations.	The other sessions were very challenging. The combination of a fast pace, challenging concepts, and numerous technical problems (problems getting modules to work with Jupiter Notebook) left me doubting my ability to perform any of the tasks covered in code lab.	N/A
Albertina Walker Hughey	Thinking about what I wanted my digital edition to actually accomplish excited me most.	The most challenging thing for me today was imagining-- from where I now sit--how to build extensibility and interoperability into my digital edition. But I'm not going to stress over it yet; the Institute will introduce technologies in the third week for building APIs that facilitate both.	I can't think of anything we accomplished today that won't influence my work and my teaching.
Andrea Nichols	Theory of Edition	Alexandria installation / Docker. Never got it installed.	Tokenization and Normalization
Christopher Ohge	Tokenization	Got to Hyper-V installation but even with enabling it and restarting, it never stayed enabled, so who knows how to get beyond that step.	Probably digital collation.
Nicola Reggiani	The possibilities of using Collatex to produce collations.	Tokenization instructions and translating them for my own work in XML, rather than text files or direct string of text (which I could get to work)	Probably the non-hierarchical text models (LMNL, TAG), which can be fruitfully flanked to XML for a better understanding and representation of the text.
	The other different ways of modelling and visualizing a text	I found XPath easier to figure out how to do it, with Tara's amazing help	
		Tokenization & normalization.	
		Tokenization and normalization, which require a lot of Python knowledge to be deployed properly	
		I think we were a bit out of sync with the material today, and it was a bit of a struggle to play around with different things, but I think the day passed too quickly from one skillset to another so there was a risk of feeling overwhelmed.	
		I could follow the work on tokenization and normalization, and appreciated the context of the gothenburg model, but I think it might've been preferable to, say, stick with tokenization for the afternoon, cover the different approaches and have us get it to work with our texts so we could consolidate our knowledge, before moving on to normalization.	
		I worry that a lot of people were left behind in the work today, from what I've heard in general conversation. I think the instructors either need to go more slowly through the material, or explain clearly what we need to take away from each sub-section (so help people understand whether they should try to be able to write python like that used to tokenize the earlier script, or if not, how else to think about the exercise). This happened a bit, with LMNL for example, but could happen more. This was compounded by difficulties installing the tech, which is hard to avoid or predict, but there may be slight changes in delivery that would help ensure that we don't lose people.	
		I think a lot of people would appreciate more short periods of group work, to check knowledge and allow us to help each other (so for example the xpath questions yesterday would have been great to do in say groups of three). Or even put us into four larger groups so we can turn to each other and take 5 minutes to reflect on what was covered in the previous hour, perhaps with an instructor.	N/A
Elizabeth Williamson	N/A		
<p>Additional comments</p> <p>I'm not sure if I'm a good informant, but I must admit was dangerously close to getting out of hand. I sympathize with the challenge of teaching these materials. And it is indeed impressive to see how this work can be done. Yet the method of teaching almost became incomprehensible to me at times. Again I thought there was an information overload with insufficient explanation. The effect of this method, in my opinion, is to reinforce the divide in digital humanities between the programmers &amp; non-programmers. It becomes intimidating for a novice programmer to look at such complex code &amp; not understand a bit of it, &amp; the instructor acts as if it is not worth explaining. In fact, I have heard the following phrase multiple times this past week &amp; a half -- "there's this thing called Google." Well, I tend to think that Google ought not to replace teachers. Otherwise we all could have had Google teach us Python rather than have an NEH Institute. This is all to say that it would be good if the instructors slowed things down a bit &amp; more clearly explained their concepts instead of throwing waves of advanced stuff at us quickly. Of course I appreciate the challenge task of the instructors, &amp; their talents. It is really impressive. And I'm still learning a lot.</p> <p>Perhaps it would be useful to have additional readings about editions available ahead of time to help everyone be thinking of all the varieties (some are in field of linguistics, others in modern while others premodern literary, some history, etc.). The readings provided are a good start too.</p>			

## Week 2, Day 3 (PhilCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu	LOD, I have worked with it before, but now I feel that I am getting a deeper understanding. I think it was also interesting to hear about the controversy surrounding LOD and the Semantic Web	Normalization. I have to admit that I was lost in this lecture. It was going very fast and I missed a few steps, and then lost track The same thing: setting us to work on tokenizing our texts—it was daunting and time-consuming to write a new script or even to figure out where to begin, and I'm really unhappy with the many errors I'm generating on simple things. Probably I'm just tired, and I often feel frustrated about spinning my wheels during workshops, but I wish I were making more progress during code labs. I'm doing much better with writing and rewriting the code in the quiet in the evening, so this is heartening.	LOD  Review of the multiple ways to approach tokenization with Python is really helpful, as well as discussion of how best to work with Python over XML, at command line and in the Jupyter notebook. I hope we'll be discussing other options like PyCharm.
Elisa Beshero-Bondar	Setting us to work on tokenizing our texts was the most exciting activity today.		
Ming Yeung Cheung	Finding out open data sites for linking personal names.	Understanding RDF that the extra review time at the end of class helped me feel more confident about it. Excellent teaching on the part of Tara, David, and the ever helpful Leif.	T-PAN transcription tools and linked open data. Again, thinking about how to take my xml documents down to a granular level to facilitate processing and analysis. I also got to thinking about the possible significance of linked open data to my project.
Roxanne Eberle	Learning how to tokenize and learning that there is more than one way to approach the problem.	I feel like I was missing some key concepts covered in D2 that were relevant for today. For the tokenization code lab I got stuck trying to strip a child tag from my XML (an emph embedded in a verse line). The advice I was given was to pick a simpler text, which was a little disheartening. My general feeling is that a lot of people are really struggling with the pace and are getting lost. I spent the free time at the end of class looking for a good tutorial on regex, Xpath, and bookmarking	
Les Harrison	Honestly, I left class today feeling woefully behind.	David's intro to collatex and XML.	N/A
Albertina Walker Hughey	Nearly everything we did to day was new, exciting, awe-inspiring.	Tokenization. Definitely tokenization.	A colleague at a neighboring institutions wants to collaborate on a study of east Texas dialects next academic year. Tokenization will figure quite prominently in that work, even though I don't anticipate using it in my in-progress critical digital edition.
Andrea Nichols	Getting more practice with Tokenization and Normalization Linked Open Data	Actually getting Tokenization and Normalization to work with my project and interests Understanding the python scripts, & seeing how it will translate to my editorial work.	Learning how to properly set up formulas for Tokenization and Normalization (Yay, Tara's help!!) Liked the UPI exercise, thinking through what to link, BEFORE creating things.
Christopher Ohge	Getting closer to understand how to tokenize texts. Tokenization: it works with Greek papyri tool And also RDF: though having been working with Papyri.info since its creation in 2010, I did not know about that linking model!		Again, collating. Linked open data will be an issue, but not clear about that right now.
Nicola Reggiani		XML Tokenization & XML in Python: I'll need more training with Xpath, probably...	Again, likely tokenization (and normalization).

### Additional comments

Did we cover text as graph? Having trouble recalling this and there are no links to the presentation materials on the GitHub week2\_day3 plan.  
Note: Broken links are starting to accumulate on some of the git pages. The links to feedback were broken on all of the daily plan pages — I had to go back up to the directory in Github and search for the pages.  
Today was an improvement on the previous day. The instructors were wise to spend more time on review.  
I thought that Tara's opening and closing sessions on Wednesday were really clear and useful, and helped to ensure that everyone had a good grasp of what we doing regarding tokenization and normalization. I would've liked to have a more in-depth explanation of Linked Open Data, as I've been introduced to it before, eg at conferences, but I would like to learn more about any possible applications in my own work or field. The instructors are doing well at dealing with various download and installation problems!  
  
The html pages for the instructor presentations aren't being hyperlinked into the daily plan as well. Making it harder to follow along by having to do more steps. I hope they get added in eventually, so that I can use the website as a reference for teaching others or finding things again.

## Week 2, Day 4 (PhilCamp)

	Comments	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Ming Yeung Cheung		Stylometry	Xquery	Stylometry and Xquery I get the feeling much of it will; I just need to tease things out a bit. I want to understand what new knowledge sets offer my work and the work of others at my institution.
Albertina Walker Hughey	I had a hard time getting the Exist software for XQuery working (had to delete and restart and reinstall). I can get the link open and input commands. Not sure if I have the proper database pulled. Not sure about how to access 'homepage' that lists all the help features and libraries.	The focus on the science behind what we're doing with these new technologies fascinated me most today.	We moved so rapidly through material that I had difficulty connecting it all; will review and research topics covered in my notes once I return home.	
Andrea Nichols	Loved Mike's sessions. Broken down well, clearly explained and modeled. Good concrete examples. Never heard of that software but thinking of ways to use it. Is there a user guide for it, so all the option buttons can be used without simply playing for hours to figure it out? Very good combination of topics. It was good to see an understandable overview of authenticity criticism combined with some rudimentary (& practical) exercises in using stylometry with R. It was also very reassuring to gain a better understanding of Collatex, & to hear Ronald tell us that it will take a while to understand all these programming languages.	N/A	N/A	N/A
Christopher Ohge	I'm also very excited about managing xml files with eXist database. Very useful for my work.	N/A	N/A	N/A Likely both text statistical analysis (for authorship evaluation of fragmentary papyrus texts) and XQuery (for developing advanced search options in the XML source of papyrus editions)
Nicola Reggiani	N/A	Text analytics and authorship evaluation	XQuery and related stuff	

## Week 2, Day 5 (PhilCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Ming Yeung Cheung	Stylometry and data visualization for Python	Xquery I continue to struggle with the quick pace with which we shift languages, programs, emphasis, etc. However, I deeply appreciate the review sessions that we've returned to at the end of the day.	Stylometry and data visualization
Roxanne Eberle	I've really enjoyed Mike's presentations; the information he presented resonates with the kinds of data/textual analysis that I'm interested in doing.	Nothe Institute.--By now, I'm starting to generate ideas for my edition based on what I'm learning. For better or worse, I'm compelled to document these ideas as they rise up while learning more about (and applying in code lab) these new technology. Thus, managing my thoughts, ideas, and framing questions while simultaneously taking-in this incredibly stimulating content presented the great challenge for me today.	In addition to Mike's presentations, I've been very interested in the challenges of xPath and xQuery, two aspects of DH that I absolutely most master.
Albertina Walker Hughey	The pipeline protocol.  LOVED R Command stylo stuff Starting to realize that I am remembering stuff so I can read error messages and figure out what to start doing to solve it in Python or Command Line. Probably not as fast as an instructor, but it's starting to stick. Yay!! Still got lots to learn though	Still need to work on Xquery	(Let me clear my thoughts-stream and get back to you on this one.)
Andrea Nichols			Modeling XML data in R-Stylo and using XQuery seem the most connected so far. Helps think about how to set up my XML tagging.
Nicola Reggiani	Xquery	The cluster and principal components analysis of texts I had some simple errors with eXist db that meant that I couldn't easily make progress during the session proper, but I really appreciated the one-to-one help I received during the break. I'll like to do some more hands-on work with this, and also learn more about the program more generally.	XQuery, again; likely PCA as well for deep analysis of papyrus texts
Elizabeth Williamson	I'm excited about using eXist db.		I think eXist db could be really useful for my work.

Additional comments  
Mike's sessions were very clear and interesting. He helpfully explained both what we were doing and why, and the roadmap for the session.  
It was nice to have a review of Python with Mike, as he's good at explaining things, making that file of examples valuable to have since I could write comments in the cells as we went along. Plus, having that time was useful for solving problems, and in case we had had to skip something earlier in institute, it's good that time was allotted to cover it.

# Week 3, Day 1 (PubCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu	CETelcean was very interesting. I struggled to understand the lecture and how it works, but I would like to investigate it further in the future	Flask. I was quite lost in that lecture	Probably CETelcean, as I think it may have applications for some of my classes. Possibly CETelcean and likely Flask. I'm also definitely interested in experimenting more with Python for comparing authors, tracing lines of influence, and comparing the "voice" of an author in multiple genres (e.g. across prose fiction and plays) provided I have a large enough data set to provide statistically meaningful results.
Elisa Beshero-Bondar Ming Yeung Chung	Reviewing and learning new ways to use Python for clustering stylistic analysis of texts. Also learning how to work with Flask. Exploring Texts using Vector Space Models	Trying to follow the slightly conflicting instructions re intro to Flask: (e.g.: Write a file and save it in your GitHub fork, when that file already exists deep at another level in the fork. Several of us sort of "lost our place" there. We would have been better off to read instructions only rather than listen to the spoken instructions. Flask	Exploring Texts using Vector Space Models I continue to think that stylometry could be important but I also know that APIs are in my future. I was also interested in the Flask material. The concept behind CETelcean--i.e., pulling together components from existing protocols to facilitate publication of my edition; the use of Flask as a dynamic content server; and the idea of using of xistdb to transform my XML and query output into my API and webpage. (I'm still working on deploying Flask and xistdb consistently.)
Roxanne Eberle	I thought today was an extremely rewarding day. Every section was engaging.	Remembering math terms	Mike's work in stylometry and maybe Flask
Albertina Walker Hughey	CETelcean and its possibilities. Mike's sessions I can follow well, so I feel like I learned that topic (the reasons AND the coding)	Keeping up w/ code lab executions because several shortcuts and Regex aren't yet part of my muscle memory. Repetitive practice should fix this situation. Not sure why spend so much time on CETelcean (it seems useful, yes).	Probably bag-of-words; certainly stylometry & CETelcean. CETelcean (for personalizing the XML sources of papyrus texts) and GitHub publishing
Andrea Nichols	Going under the hood with bag-of-words analyses. Also excited about using CETelcean.	The final session on flask.	
Christopher Ohge		Stylometry and Flask It would've been really useful to have at the start of the session something like a brief roadmap for the session, the end goal articulated, and importantly tell us where we need to be - eg we're going to be using python in the command line, you can do this by pulling from the git repo and then executing the existing python scripts, or we should be in the jupyter notebooks in folder x, or whatever. When we get this in sessions it's really helpful, because when we're changing between tech and approaches and instructors frequently, you can easily get lost at the start and then you miss vital info.	
Nicola Reggiani	CETelcean and GitHub publishing	One thing I wasn't 100% sure on this evening: I'm confused about whether the words in the render_template were calling html files or whether the content was just strings. See example in flask-index-template-dyn.py return render_template("index-dyn.html", title="--template index.html", header="index.html", paragraph="This is index.html with dynamic contents in response to a request for / (ROOT) in flask-index-template-dyn.py") -- is this the same as header="apples and bananas" ?	
Elizabeth Williamson	I think I understood Leif's session on Flask more after the class, when I went through reading the text and looking up things like web template systems. It was useful to have the schedule text there, as always. This does look really interesting and useful.		I'm looking forward to seeing how Flask works with the API work planned later in the week.
<p>Additional comments</p> <p>The first two sections were excellent; Mike is very good at explaining difficult subjects. CETelcean looks like a very promising tool, but I would like to hear more about it. I wonder if we could return to it during a review session?</p> <p>It would also be good if we could review the material on Flask. It was not presented very clearly, &amp; it didn't seem occasion enough background material.</p>			

## Week 3, Day 2 (PubCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu	LOD. I feel like I'm getting a pretty good grasp on the relationships among XML, XSLT, RDF, etc. Learning how I might implement eXist-db more fully and directly than I have been in the past into the web delivery of my projects. Learning about eXist apps.	the eXist lesson was difficult to follow	LOD and rdf
Elisa Beshero-Bondar Ming Yeung Cheung	N/A	Understanding how and why to implement LOD, but this was also an enlightening topic. Building API in eXist-db and XQuery	Web implementations and applications of eXist-db. Also, the writing of XSLT to generate RDF--seeing examples of that from Hugh was really helpful.
Paul Hackett	learning how to configure eXist	Understanding the utility and mechanism for constructing RDF	RDF
Albertina Walker Hughey	Pulling things together logistically and receiving feedback from faculty and cohort members.	Nothing related to the Institute. (Stayed up late drafting ideas for reducing the scale of edition as originally conceived, deciding to deploy it as two sites. Completed dazzled by possibilities...) Plugging your edition into a graph (RDF / SPARQL) Getting an app to open in XQuery	eXist app & webserver
Andrea Nichols Nicola Reggiani	Maybe making apps in XQuery? thinking through how all our new software skills fit together RDF/SPARQL	For both of those sessions, it was a matter of making software work. I got the application steps sorted out later, with help from Les (and I helped Francesca in turn, who was having the same issue) eXist and XQuery webapps	The use of elements and applications from multiple programming languages to accomplish project-specific outcomes, and distinctions between server-side decisions/tools/protocols and the client-side ones.
<p>Additional comments</p> <p>Idea: In future NEH DH Institutes, could just have one teaching unit per week be for showing the 'teacher's best resources' tools / software / help guides / videos / blogs etc instead of teaching some things for an hour and half. That way, time could be spent more on the basic underpinning languages in Python, XPath, XQuery, etc. that many things seem to draw upon. This would also free up some time for sorting out software issues...have a "Troubleshooting Lab" time built in one day a week, to install everything needed that week, and sort out issues without having people get behind because they can't follow the teaching unit directions while troubleshooting.</p> <p>Idea: Maybe in future NEH DH Institutes, start with some examples of digital editions, and show all the features and what it does, then break that apart saying "In the following weeks we will learn..." and tie the software and topics into each of the features, so that as we do things we can mentally visualize how each piece builds up into the final product.</p> <p>I figured out why I like Mike's sessions and grasp them (even though it's been years since I was in statistics or algebra/geometry). He sets up small steps and gives detailed explanation on not only how it works, but making us think about what that will do to the data in concrete examples. Also, Mike's method of printing out each row of code in Python and making sure we understand it, and even after explaining, going back over it, makes it so much clearer. Gives my mind time to grasp AND retain (by taking notes...if I don't have time to take notes in my own words, then I haven't grasped it). As with Tara's "read the error message" and Leif's "read the documentation" mantras, Mike's method of "print" in Python will stick with me on how to figure out the code.</p>			

## Week 3, Day 3 (PubCamp)

	What excited you most about what we did today?	What did you find most challenging of what we did today?	What topic addressed today is most likely to influence your future work?
Marie-Claire Beaulieu	I was very interested in hearing everyone else's talks. I think it would have been useful if we did this earlier in the Institute so as to have the different research questions in mind.	The TEI stylesheets were difficult, also I had lots of trouble with Docker, so it was hard to follow the lesson nothing especially challenging today, but attempting to repair the TEI stylesheets in pointing to dependent css proved difficult--I didn't get that working and gave up trying--probably just had to keep tinkering with my Docker installation and its distinctive file paths. It's a humbling thought how difficult those Stylesheets are for the TEI to maintain when things break.	I found it useful to articulate my project to this audience, and I think the feedback received will be very useful.
Elisa Beshero-Bondar Ming Yeung Cheung	listening to Institute participants deliver lightning talks on their fascinating projects.	N/A	discussions of collation strategies.
Roxanne Eberle	It is hard not to say the lightning talks. I really enjoyed seeing the range of work that my colleagues are engaged in.	I had some struggles with getting the command line conversion to pdf to work but I can see how useful that ability will be (once I conquer the various syntactical and technological hurdles that I still have).	not sure
Albertina Walker Hughey	The lightninging talks; seeing, hearing, and appreciating cohort members' in-progress digital editions.	I felt comfortable today with code lab and content reiterations. Appreciating the conceptual flow of what we're doing here. No challenges today.	I was so impressed with the elegance and attention to user interaction demonstrated by Isabella and Christopher's projects, as well as the very interesting visualizations generated by Digital Ricoeur.
Andrea Nichols	EEBO TEI!!!! Yay! I live in EEBO.	Couldn't get the transform to pdf step to work during TEI	Geo-mapping technologies.
Nicola Reggiani	Discussion and comparison with the others' projects	RDF / SPARQL	It was nice to hear other projects (but only for 5 min, not 10 min, given attention span at this point in the seminar) to get to see what people are producing. Had good discussion with seat-row mates on "what is an edition?" and swapping readings and resources (that continued on Friday with new seat-mates). Perhaps Text as Graph, to test different ways of approaching and visualizing text

## Week 3, Day 4 (PubCamp)

Marie-Claire Beaulieu Ming Yeung Cheung Albertina Walker Hughey	What excited you most about what we did today? I was excited about the morning sessions on APIs. I felt that things were explained simply and in a way I could follow with success to demystify this concept Able to write API The post-class eXistdb faculty-led coding session.	What did you find most challenging of what we did today? Unfortunately I had to miss the afternoon sessions, but that seemed challenging! How exactly APIs work Institute day was fine. No notable challenges today. Understanding what APIs were beyond the literal description...how to visualize them on top of the other programs we had learned	What topic addressed today is most likely to influence your future work? I thought the resources gathered by the organizers on APIs and other topics will be extremely helpful in the future when I need to return to various topics. Some API knowledge The Pipeline. All of them (API, licensing, and eXist-db evening session)
Andrea Nichols	Pretty much every session was helpful and interesting and exciting. Things really started stitching together		
Nicola Reggiani	API designing	API construction	Again, the API thing - I would definitely do my best to build an API for linguistic variation analysis in Greek papyri!

## Week 3, Day 5 (PubCamp)

Marie-Claire Beaulieu	What excited you most about what we did today? I enjoyed the discussion of publication and sustainability, as it is a crucial issue in DH	What did you find most challenging of what we did today? The further discussion of TAG and Alexandria, but I am curious about it	What topic addressed today is most likely to influence your future work? Graphs and visualization. I am excited to experiment with it! I'll need to continue grappling with the question of visualizations and their potential role in my edition. I'd like to use them strategically as a way of illuminating my research questions and also employ them as a way for users to get more out of my edition.
Roxanne Eberle	I do really feel as though the seminar has met its desired outcomes: I'm excited about continuing to explore the multiple digital affordances that are out there and confident that I'll be able to produce a better critical edition.	The most challenging aspects continue to be technological but as this seminar concludes I have greater faith that I'll be able to meet those challenges. Nothing that won't work itself by developing my edition's infrastructure, testing it out, and publishing it in order to review, re-configure as needed, relaunch, and monitor use/feedback. Data visualization (some Python code worked, some didn't) The Stemmaweb software under development seems very interesting and useful if I ever need that. But I don't do stemmas at this point	
Albertina Walker Hughey	Putting it all together in multiple scenarios recursively. Publishing, deployment, acronyms etc (stitching things together more and building and hearing others' experiences)		Everything we addressed today will influence my future work. This Institute's a game changer.
Andrea Nichols	Being able to help seat-row mates sort out file finding and loading issues		Publishing, deployment, acronyms etc

## Week 3, Day 6 (Pedagogical retrospective)

Andrea Nichols	What excited you most about what we did today? Discussing how it was organized to work & what we learned, what we thought coming in, seeing how the methods & modeling changed our viewpoints, how much we learned and how we feel empowered to make our editions instead of <i>copying</i> someone, and to <i>teach</i> others	What did you find most challenging of what we did today? Waking up after a long week and rainy night, haha	What topic addressed today is most likely to influence your future work?  All of it. Hearing ideas and concerns on teaching, reflection on what we have learned, how we feel empowered, how we will make and improve our projects due to having attended, etc.
<p><b>Additional comments</b></p> <p>...instructors and assistants. For the instructors, modeling the patience and "roll with it" diagnoses of issues as we learned new things and downloaded software. It is just as important to know the tools, as it is to "read the error message", "print the Python code" to see what it does, Google an answer (but knowing better how to word that and evaluate the type of answer we may see in the results page), and read the manual so that we don't quit a few hours into attempting something. For the assistants, staff at Pittsburgh, and organizers / instructors (I'm sure a vast network was involved), it was so helpful to have someone meeting us to help us know where to go upon arrival, having the ID cards prepared, wonderful dorm rooms, a meal plan, coffee during breaks, getting basic admin questions answered that arose during our stay, and the close proximity of all the places we needed (classrooms, food, dorm, laundry room, and drug store / coffee shops). I have made new friends, new connections, and learned about more projects. I feel more empowered to tackle what I* want an edition to do, instead of copying someone else because I have little ability or confidence to hammer on the code, or handing off my XML to a programmer and hoping for the best.</p> <p>I could see where, even with moving to another state within the next few days, that faculty in NC who know me and already want me to come by their classes, may have me not only teach about early modern print and Tudor England, but also about DH. Thank you for the phenomenal opportunity to learn and 'spread the Gospel' as I go forth in my career (wherever that may take me).</p>			